

**CLONING, EXPRESSION, CRYSTALLIZATION AND
PRELIMINARY X-RAY CRYSTALLOGRAPHIC DATA
OF THE *Pyrococcus furiosus* THERMOSTABLE
DNA POLYMERASE**

Anantasak Loonchanta

**A Thesis Submitted in Partial Fulfillment of the Requirements for the
Degree of Master of Science in Biotechnology
Suranaree University of Technology
Academic Year 2006**

การโคลน ผสม ตกผลึกและข้อมูลเบื้องต้นของการ X-ray ผลึก
ดีเอ็นเอโพลีเมอร์สทนร้อนจาก *Pyrococcus furiosus*

นายอนันต์ศักดิ์ คุณจันทา

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาเทคโนโลยีชีวภาพ
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2549

**CLONING, EXPRESSION, CRYSTALLIZATION AND
PRELIMINARY X-RAY CRYSTALLOGRAPHIC DATA OF THE
Pyrococcus furiosus THERMOSTABLE DNA POLYMERASE**

Suranaree University of Technology has approved this thesis submitted in
partial fulfillment of the requirement for a Master's Degree.

Thesis Examining Committee

(Asst. Prof. Dr. Chokchai Wanapu)

Chairperson

(Asst. Prof. Dr. Mariena Ketudat-Cairns)

Member (Thesis Advisor)

(Assoc. Prof. Dr. James Ketudat-Cairns)

Member

(Assoc. Prof. Dr. Saowanee Rattanaphani)

Vice Rector for Academic Affairs

(Asst. Prof. Dr. Suwayd Ningsanond)

Dean of Institute of Agricultural Technology

อนันต์ศักดิ์ ลุนจันทา : การโคลน ผลิต ตกผลึกและข้อมูลเบื้องต้นของการ X-ray ผลึก
ของดีเอ็นเอโพลิเมอร์เรสทนร้อนจาก *Pyrococcus furiosus* (CLONING, EXPRES-
SION, CRYSTALLIZATION AND PRELIMINARY X-RAY CRYSTALL-
OGRAPHIC DATA OF THE *Pyrococcus furiosus* THERMOSTABLE DNA
POLYMERASE) อาจารย์ที่ปรึกษา : ผู้ช่วยศาสตราจารย์ ดร.มารินา เกตุทัต-คาร์นส์,
94 หน้า.

เอนไซม์ดีเอ็นเอโพลิเมอร์เรสทนร้อนจาก *Pyrococcus furiosus* หรือ *Pfu* ดีเอ็นเอโพลิ-
เมอร์เรส จัดอยู่ในกลุ่มเอนไซม์ดีเอ็นเอโพลิเมอร์เรส family B เป็นเอนไซม์ที่มีความสามารถในการ
เติมเบสนิวคลีโอไทด์ในปฏิกิริยา PCR ได้แม่นยำที่สุดในสภาวะที่มีอุณหภูมิสูง ในการศึกษากลไก
และการควบคุมการเติมนิวคลีโอไทด์หรือความสามารถในการทนความร้อนสูงของเอนไซม์ในเชิง
โครงสร้างโปรตีน จำเป็นจะต้องมีโครงสร้างของเอนไซม์นี้ อย่างไรก็ตาม จากการศึกษาค้นคว้า
พบว่า ยังไม่มีการวิจัยเพื่อหาโครงสร้างของเอนไซม์นี้ โดยการศึกษานี้มีวัตถุประสงค์เพื่อ โคลน
ผลิต ทำโปรตีนให้บริสุทธิ์ วัดกิจกรรมของเอนไซม์ ตกผลึก และ นำเสนอข้อมูลทางด้านเอ็กซ์เรย์
เบื้องต้นของเอนไซม์ *Pfu* ดีเอ็นเอโพลิเมอร์เรส

ในผลการศึกษา ได้มีการโคลนยีนดีเอ็นเอโพลิเมอร์เรสจากจีโนมของ *P. furiosus* เข้าสู่
pSY5 พลาสมิด และผลิตโปรตีนใน *Escherichia coli* Rosetta (DE3) pLysS ในรูปแบบที่เป็น
His₈-tagged *Pfu* ดีเอ็นเอโพลิเมอร์เรส โดยมีความสามารถในการผลิต 38 มิลลิกรัมต่อลิตร
ได้มีการทำ โปรตีนให้บริสุทธิ์และวัดกิจกรรมของเอนไซม์ ซึ่งพบว่ามี relative activity 30,000
U ต่อมิลลิกรัมโปรตีน และยังพบอีกว่าเอนไซม์ที่ผลิตนี้ยังมีความคงทนต่อสภาพความร้อนที่ 97.5
°C เป็นเวลานานถึง 23 ชั่วโมงและมีความสามารถในการทำ PCR ได้สูงกว่าเอนไซม์ที่มีขาย
ตามท้องตลาด อย่างไรก็ตาม เอนไซม์ได้มีการสูญเสียความสามารถในการทำ PCR บ้างเมื่ออยู่ใน
สภาพความร้อนดังกล่าวข้างต้น

ในการตกผลึกโปรตีนนั้น ได้มีการกระบวนการทำแตกต่างกันสองกระบวนการ คือ ไม่มี
การให้ความร้อนโปรตีนก่อนการตกผลึกและมีการให้ความร้อนก่อนการตกผลึก จากผลการศึกษา
พบว่า โปรตีนชนิดไม่ได้ให้ความร้อนมีผลึกเดียวเกิดขึ้นในสภาวะที่มี 12% (w/v) PEG1000,
200 mM ammonium phosphate (monobasic) และสามารถหักเหเอ็กซ์เรย์ ได้เพียงที่ 4
อังสตรอม เท่านั้น ส่วนโปรตีนชนิดให้ความร้อน ซึ่งมีผลึกเดียวเกิดขึ้นในสภาวะ ที่มี 10% w/v
PEG8000, 100 mM sodium acetate, and 50 mM magnesium acetate สามารถหักเหเอ็กซ์เรย์ได้ที่ 3 อังสตรอม โดยมี unit-cell parameter เป็น $a = 91.9 \text{ \AA}$, $b = 126.8 \text{ \AA}$, $c = 88.4 \text{ \AA}$, $\alpha = 90.0^\circ$, $\beta = 109.1^\circ$, and $\gamma = 90.0^\circ$ และ เป็นผลึกชนิด monoclinic

space group C2 โดยมี Matthew's coefficient เท่ากับ $2.64 \text{ \AA}^3 \text{Da}^{-1}$ และมี solvent content เท่ากับ 53.5% (v/v)

โครงสร้างโดยรวมของเอนไซม์ *Pfu* ดีเอ็นเอโพลีเมอเรสในเบื้องต้นประกอบด้วยโดเมนและซับโดเมนต่าง ๆ ดังต่อไปนี้ คือ โดเมน N-terminal, Exonuclease และ Polymerase ซึ่งโดเมน Polymerase นี้ประกอบด้วย 2 ซับโดเมน คือ ซับโดเมน Palm และ Fingers โดยที่ไม่สามารถเห็นโครงสร้างของโดเมน Thumb นอกจากนี้ยังสามารถเห็นพันธะ disulfide Cys452-Cys466 และ Cys530-Cys533 ที่เป็นจุดเชื่อมต่อระหว่างซับโดเมน Palm และ Fingers

ANANTASAK LOONCHANTA : CLONING, EXPRESSION, CRYSTAL-
LIZATION AND PRELIMINARY X-RAY CRYSTALLOGRAPHIC DATA
OF THE *Pyrococcus furiosus* DNA POLYMERASE. THESIS ADVISOR :
ASST. PROF. MARIENA KETUDAT-CAIRNS, Ph.D. 94 PP.

PFU DNA POLYMERASE/PYROCOCCUS FURIOSUS/CRYSTALLIZATION/
X-RAY DATA

The *Pyrococcus furiosus* thermostable DNA polymerase or *Pfu* DNA polymerase is structurally homologous to the family B DNA polymerases. It has been shown to have the highest fidelity, introducing the lowest amplification errors in PCR products. Nevertheless, the structure of *Pfu* DNA polymerase is unknown. Understanding of the structural mechanisms and control of its high fidelity, as well as its thermostability is therefore limited. The objectives of this study were to clone, express, and purify *Pfu* DNA polymerase, test its activity, crystallize it and generate the preliminary X-ray crystallographic data.

The DNA polymerase gene was cloned from *P. furiosus* genomic DNA into the pSY5 plasmid vector, and expressed in *Escherichia coli* strain Rosetta (DE3) pLysS as the His₈-tagged *Pfu* DNA polymerase with yield of 38 mg/L culture. The proteins was purified and tested for its activity. The relative activity is 30,000 U/mg protein. The His₈-tagged *Pfu* DNA polymerase was still stable when incubated at 97.5 °C for 23 h and has higher PCR efficiency than commercial *Pfu* DNA polymerase. Some loss of PCR efficiency occurred in the heat-treated protein.

Single crystals of non-heated His₈-tagged *Pfu* DNA polymerase were obtained from a condition containing 12% (w/v) PEG1000, 200 mM ammonium

phosphate (monobasic) and diffracted to a resolution limit of only 4 Å. On the other hand, single crystals of heated His₈-tagged *Pfu* DNA polymerase were obtained from 10% w/v PEG8000, 100 mM sodium acetate, and 50 mM magnesium acetate. The crystal diffracted to a resolution limit of 3 Å. The unit-cell parameters were determined as $a = 91.9$ Å, $b = 126.8$ Å, $c = 88.4$ Å, $\alpha = 90.0^\circ$, $\beta = 109.1^\circ$, and $\gamma = 90.0^\circ$ with the monoclinic space group of C2, which gave Matthew's coefficient of $2.64 \text{ Å}^3\text{Da}^{-1}$ and a solvent content of 53.5% (v/v).

The preliminary overall structure is basically composed of an N-terminal domain, Exonuclease domain, and Polymerase domain including the Palm and Fingers subdomains. The Thumb domain is absent in the structure. Two disulfide bonds are found in the connection site between the Palm and Fingers subdomains of the structure: Cys452-Cys466 and Cys530-Cys533.

School of Biotechnology

Academic Year 2006

Student's Signature_____

Advisor's Signature_____

ACKNOWLEDGEMENTS

This work was conducted mainly at Bob Robinson laboratory (BR Lab), Institute of Molecular and Cell Biology (IMCB), Singapore. Suranaree University of Technology and IMCB funds acknowledge the funding bodies for financial support.

I express my deepest sense of gratitude to my advisor Asst. Prof. Dr. Mariena Ketudat-Cairns for giving me full freedom in my work, while at the same time she was always cheerful and supportive. This work could never be initiated without her.

I express my great appreciation to my co-advisor Assoc. Prof. Dr. Robert Robinson for allowing me to work and utilize the fascinating facilities freely in his laboratory so that the work can be finished sooner.

My deepest appreciation is expressed to my greatest teacher Mr. Sakesit Chumnarnsilpa for his encouragement, valuable suggestion, constructive idea, and criticism of the research. This thesis could never be done without him.

I would like to thank Assoc. Prof. Dr. James Ketudat-Cairns for his assessment and suggestion on the thesis report and also Asst. Prof. Dr. Chokchai Wanapu for his partial correction of the thesis.

I thank all my friends and colleague for their valuable support.

Above all I thank my dear family, whose constant support and encouragement has made me a successful person.

Anantasak Loonchanta

CONTENTS

	Page
ABSTRACT IN THAI.....	I
ABSTRACTS IN ENGLISH.....	III
ACKNOWLEDGEMENTS.....	V
CONTENTS.....	VI
LIST OF TABLES.....	X
LIST OF FIGURES.....	XI
LIST OF ABBREVIATIONS.....	XIII
 CHAPTER	
I. INTRODUCTION.....	1
1.1 The problem.....	2
1.2 Objectives.....	3
1.3 Scope.....	3
1.4 Benefits.....	3
1.5 Conceptual framework.....	4
II. LITERATURE REVIEW.....	5
2.1 Thermostable DNA polymerase.....	5
2.2 Crystal structure of archaeal family B DNA polymerase.....	8
2.3 Protein crystallization.....	11
2.3.1 Principles of protein crystallization.....	11
2.3.2 Protein crystallization methods.....	12

CONTENTS (Continued)

	Page
2.3.2.1 Vapor diffusion.....	12
2.3.2.2 Dialysis.....	13
2.3.2.3 Batch technique.....	13
2.3.2.4 Microbatch.....	14
2.3.3 Symmetry in protein and macromolecular crystals.....	14
2.4 Determination of the crystal structure of proteins.....	16
2.4.1 Data collection.....	16
2.4.1.1 Scattering from crystals.....	16
2.4.2 Phasing.....	18
2.4.3 Model building.....	20
2.4.4 Model refinement and validation.....	21
2.5 Quality indicator for the protein structure.....	22
2.5.1 The R factor.....	22
2.5.2 The free R factor.....	23
III. MATERIALS AND METHODS.....	25
3.1 Materials.....	25
3.1.1 Bacterial strains.....	25
3.1.2 Expression plasmid vector.....	25
3.1.3 Enzymes.....	25
3.1.4 Reagents and kits.....	26
3.2 Apparatuses.....	26

CONTENTS (Continued)

	Page
3.3 Methods.....	26
3.3.1 Cloning of the DNA polymerase gene from <i>P. furiosus</i>	26
3.3.2 Expression of recombinant His ₈ -tagged <i>Pfu</i> DNA polymerase.....	28
3.3.3 Purification of recombinant His ₈ -tagged <i>Pfu</i> DNA polymerase.....	28
3.3.4 Relative DNA polymerase activity assay.....	30
3.3.5 Determination of His ₈ -tagged <i>Pfu</i> DNA polymerase thermostability.....	31
3.3.6 Examination of His ₈ -tagged <i>Pfu</i> DNA polymerase PCR efficiency.....	31
3.3.7 Crystallization of His ₈ -tagged <i>Pfu</i> DNA polymerase.....	32
3.3.8 Data collection and processing from the His ₈ -tagged <i>Pfu</i> DNA polymerase crystals	33
3.3.9 Preliminary structure determination of the His ₈ -tagged <i>Pfu</i> DNA polymerase.....	34
3.3.9.1 Phasing.....	34
3.3.9.2 Model building and refinement.....	35
IV. RESULTS AND DISCUSSIONS.....	36
4.1 Cloning of the DNA polymerase gene from <i>P. furiosus</i>	36
4.2 Expression of recombinant His ₈ -tagged <i>Pfu</i> DNA polymerase.....	36
4.3 Purification of recombinant His ₈ -tagged <i>Pfu</i> DNA polymerase.....	39
4.4 Relative DNA polymerase activity assay.....	41

CONTENTS (Continued)

	Page
4.5 The determination of His ₈ -tagged <i>Pfu</i> DNA polymerase thermostability...	42
4.6 The examination of His ₈ -tagged <i>Pfu</i> DNA polymerase PCR efficiency.....	43
4.7 Crystallization of His ₈ -tagged <i>Pfu</i> DNA polymerase.....	47
4.8 Diffraction data.....	53
4.9 Preliminary model.....	53
V. CONCLUSIONS.....	57
REFERENCES.....	60
APPENDICES.....	67
APPENDIX A Sequence of cloned His ₈ -tagged <i>Pfu</i> DNA polymerase.....	68
APPENDIX B Determination of relative polymerase activity.....	71
APPENDIX C Example of crystals from crystallization screening.....	74
APPENDIX D Crystallization optimization conditions.....	77
APPENDIX E Programs used in this study.....	81
APPENDIX F Crystallographic theory.....	88
BIBLIOGRAPHY.....	94

LIST OF TABLES

Table	Page
2.1 Types of symmetry found in proteins and macromolecules.....	15
3.1 Primers used in extension efficiency assay.....	33
4.1 Hit conditions at 15 °C for crystallization of non-heated His ₈ -tagged <i>Pfu</i> DNA polymerase.....	46
4.2 Hit conditions at 15 °C for crystallization of heated His ₈ -tagged <i>Pfu</i> DNA polymerase.....	48
4.3 Data collection statistics from the heated His ₈ -tagged <i>Pfu</i> DNA polymerase crystal.....	52
1D Crystallization optimization condition OPT7 (Cryo I/II: G3).....	78
2D Crystallization optimization condition OPT9 (Ozma 1K4K: A7).....	79
3D Crystallization optimization condition OPT9 (Ozma 8K10K: G7).....	80
1F Protein crystal symmetry.....	90

LIST OF FIGURES

Figure	Page
2.1 Structure of archeal DNA polymerases.....	9
2.2 Sequence alignment of archaeal DNA polymerases.....	10
2.3 Phase diagram of crystal growth.....	12
2.4 Crystallographic data collection.....	17
4.1 PCR products generated by the use of different primers.....	37
4.2 Small scaled expression recombinant His ₈ -tagged Pfu DNA polymerase in <i>E. coli</i> Rosetta(DE3)pLysS.....	38
4.3 Small scaled expression of recombinant His ₈ -tagged Pfu DNA polymerase in <i>E. coli</i> BL21-Codonplus (DE3)-RIL.....	39
4.4 Purification profile of His ₈ -tagged <i>Pfu</i> DNA polymerase.....	40
4.5 Purity of the recombinant His ₈ -tagged <i>Pfu</i> DNA polymerase after purification by automated AKTApurification System.....	41
4.6 The relative DNA polymerase activity assay of His ₈ -tagged <i>Pfu</i> DNA Polymerase.....	42
4.7 Thermostability of His ₈ -tagged <i>Pfu</i> DNA polymerase.....	43
4.8 Comparison of PCR efficient of 5U of the recombinant His ₈ -tagged <i>Pfu</i> DNA polymerase and commercial Promega <i>Pfu</i> DNA polymerase.....	44
4.9 PCR efficient of 23 h heat-treated His ₈ -tagged <i>Pfu</i> DNA polymerase.....	45
4.10 Crystals of non-heated His ₈ -tagged <i>Pfu</i> DNA polymerase obtained from the screening plate.....	50

LIST OF FIGURES (Continued)

Figure	Page
4.11 Example of crystals of non-heated His ₈ -tagged <i>Pfu</i> DNA polymerase obtained from optimization OPT9-F6.....	51
4.12 Example of crystals of heated His ₈ -tagged <i>Pfu</i> DNA polymerase obtained from vapor diffusion: sitting drop condition HTS I: G3.....	51
4.13 Preliminary overall structure of the heated His ₈ -tagged <i>Pfu</i> DNA Polymerase.....	55
4.14 The 2 Fo - Fc electron density maps contoured at 1 σ at 3.0 Å resolution for the two disulfide bonds in the heated His ₈ -tagged <i>Pfu</i> DNA polymerase....	56
1A Sequence of cloned His ₈ -tagged <i>Pfu</i> DNA polymerase.....	69
1B The band intensities determined by Quantity One® Software, with the arbitrary unit.....	73
1C Example of crystals of non-heated His ₈ -tagged <i>Pfu</i> DNA polymerase obtained from various conditions of crystallization screening.....	75
2C Example of crystals of heated His ₈ -tagged <i>Pfu</i> DNA polymerase obtained from various conditions of crystallization screening.....	76

LIST OF ABBREVIATIONS

bp	=	Base pair
°C	=	Degree Celsius
DTT	=	Dithiothreitol
EDTA	=	Ethylenediaminetetraacetic acid
h	=	Hour
HCl	=	Hydrochloric acid
kb	=	Kilobase
µg	=	Microgram
µl	=	Microlitre
ml	=	Millilitre
mM	=	Millimolar
min	=	Minute
PAGE	=	Polyacrylamide Gel Electrophoresis
PCR	=	Polymerase Chain Reaction
PEG	=	Polyethylene Glycol
s	=	Second
SDS	=	Sodium Dodecyl Sulfate
U	=	Unit

CHAPTER I

INTRODUCTION

The structural gene encoding *Pyrococcus furiosus* DNA polymerase or *Pfu* DNA polymerase has been cloned and found to contain 2,785-bp, which consist of 775 amino acids as deduced from the DNA sequence, and has a molecular weight of 90.109 kDa. This protein is homologous to the α -like DNA polymerases (family B) represented by human DNA polymerase α and *Escherichia coli* DNA polymerase II (Uemori, *et al.*, 1993).

Polymerase Chain Reaction or PCR has become an indispensable part of modern molecular biology and molecular medicine. The use of high fidelity DNA polymerases in the PCR is essential for reducing the introduction of amplification errors in PCR products that will be cloned, sequenced, and expressed. The *Pfu* DNA polymerase has been shown to have the highest fidelity among the other thermostable DNA polymerases. The error rate of *Pfu* DNA polymerase is reported to be 7- to 10-fold lower than that of nonproofreading *Taq* DNA polymerase, and 2- to 3-fold lower than other proofreading enzymes (Lundberg, *et al.*, 1991; Eckert and Kunkel, 1991; Cline, *et al.*, 1996). Consequently, it has gained considerable attention in the field of DNA amplification and been widely used in the laboratory (Lu and Erickson, 1997).

The *Pfu* DNA polymerase possesses thermostability and the highest fidelity among other thermostable DNA polymerases so far identified, therefore, understanding the structural based hyperthermostability as well as mechanisms and

control of *Pfu* DNA polymerase proofreading activity will be a basic contribution to understanding the fundamental cell biological processes. Moreover, engineering of *Pfu* DNA polymerase in order to improve its activity can be conducted, following its structure determination. One crucial step in crystal structure determination is crystallization. The crystallization conditions that result in single crystals and high resolution diffraction data are preferable. The *Pfu* DNA polymerase protein was previously crystallized from 0.08 M ammonium sulfate, 0.05 M Nacacodylate, pH 6.5, 0.15%(v/v) NP40, 0.05%(v/v) Tween 20 and 4.5%(w/v) polyethylene glycol 6000 by the vapor-diffusion method, yielding long rod-like single crystals (Goldman, *et al.*, 1998). The crystals diffracted only beyond 4 Å on a synchrotron radiation source.

1.1 The problem

The structure of *Pfu* DNA polymerase is required in order to understand the structural based hyperthermostability, as well as the mechanisms and control of *Pfu* DNA polymerase high proofreading activity, which will be a basic contribution to understanding the fundamental cell biological processes. Moreover, engineering of *Pfu* DNA polymerase in order to improve its activity can be done, following its structure determination. However, due to the low resolution of the diffraction data obtained by Goldman, *et al.* (1998), the structure of *Pfu* DNA polymerase has never been determined. This suggested the requirement of other crystallization conditions or optimization of crystallization conditions to obtained single crystals that diffract to higher resolution.

1.2 Objectives

To achieve the aims of the thesis, the following specific objectives were investigated.

1.2.1 To clone DNA polymerase gene from *Pyrococcus furiosus*.

1.2.2 To express, purify, and test activity of *Pfu* DNA polymerase.

1.2.3 To crystallize recombinant *Pfu* DNA polymerase.

1.2.4 To elucidate the preliminary X-ray crystallographic data of *Pfu* DNA polymerase.

1.3 Scope

This study comprised cloning and expression of recombinant His₈-tagged *Pfu* DNA polymerase enzyme in *E. coli*, followed by protein purification and activity assay. Finally, crystallization, data collection and preliminary X-ray crystallographic data analysis were carried out.

1.4 Benefits

The structure of *Pfu* DNA polymerase will be of use for understanding of the mechanisms and control of its high fidelity as well as its hyperthermostability. These will be a basic contribution to understanding the fundamental cell biological processes as well. In addition, since the need for *Pfu* DNA polymerase enzyme is increasing, considerably large amounts of *Pfu* DNA polymerase have been purchased from commercial companies. Therefore, having *Pfu* DNA polymerase produced in our own laboratory can be considered as an economic benefit.

1.5 Conceptual framework

Pfu DNA polymerase used in this study is recombinantly expressed with MAEEHHHHHHHLEVLFGQGRP tag at its N-terminus. It is so called His₈-tagged *Pfu* DNA polymerase. Consequently, the protein contained 798 amino acids with a calculated molecular weight and isoelectric point (pI) of 92.865 kDa and 7.1, respectively. The extinction coefficient for $A_{280} = 0.74$ for 1 mg/ml.

The commercial Promega *Pfu* DNA polymerase, which is a natural enzyme directly purified from *Pyrococcus furiosus*, was used as a reference in the determination of relative DNA polymerase activity and PCR efficiency assay to compare with the His₈-tagged *Pfu* DNA polymerase.

The structure of *Pfu* DNA polymerase was represented by the His₈-tagged *Pfu* DNA polymerase and it is a preliminary structure. Cycles of refinement, model building and validation will be needed in order to obtain the final structure. However, this preliminary structure could be investigated globally such as the domains and loops.

CHAPTER II

LITERATURE REVIEW

2.1 Thermostable DNA polymerase

Many of the thermophilic DNA polymerases appear to be monomeric in solution with molecular masses of 80-115 kDa. They have been isolated and characterized previously from either mesophilic eubacteria or archaea sources (Klimczak, *et al.*, 1985; Rella, *et al.*, 1990; Elie, *et al.*, 1989). Corresponding to the thermal extremes of the environment from which their source species were isolated, these enzymes have thermal stabilities. Although thermal stabilities of the native proteins vary among enzymes, the range of 70-80 °C are common optimal temperatures for their polymerization activities (Rella, *et al.*, 1990; Elie, *et al.*, 1989). Considering the thermostability of these enzymes and their potential industrial applications, their structure and function relationships are of interest to researchers. Nevertheless, very low levels of most native thermostable enzymes are synthesized by the thermophilic bacteria and they are therefore cumbersome to purify.

More than 50 DNA polymerase genes have been cloned and sequenced from various organisms, including thermophiles and archaea. Their deduced amino acid sequences have been compared. On the basis of similarities in these amino acid sequences, DNA polymerases have been classified into four major groups: *Escherichia coli* DNA polymerase I (family A), DNA polymerase II (family B, α -like DNA polymerases), DNA polymerase III (family C), and others (family X) (Ito and

Braithwaite, 1991). Extremely thermostable DNA polymerases have been purified from some archaeon and the genes have been cloned (Mathur, *et al.*, 1991; Perler, *et al.*, 1992). The deduced amino acid sequences of these thermostable DNA polymerases showed that they all belong to the α family (family B).

With their integrated 3'-5' exonuclease activity that corrects errors introduced during polymerization, several thermophilic DNA polymerases, including *Pfu*, *Vent*, *deep Vent*, *Pwo*, and *Ultma* possess proofreading ability. However, they differ in their error rate, which was determined by the *lacI* PCR mutation assay described by Lundberg (1991). Briefly, the *lacIOZ α* nucleotide sequence was amplified from transgenic mouse genomic DNA containing a *lacIOZ α* . The PCR products were cleaved at unique EcoRI sites introduced at the ends of the fragments by the PCR primers, then cloned into the phage λ gt 10 vector, packaged and plated with a *lacZ Δ M15* *E. coli* host strain containing the α -complementing portion of the *lacZ* gene. A nonfunctional repressor protein is expressed when certain mutations occurred within the *lacI* gene during the PCR amplification process, generating a blue plaque when plated with top agar containing XGal. Therefore, the observed mutation frequency (frequency of phenotypic mutants) can be calculated by dividing the number of blue plaques (*LacI*⁻ mutants) by the total number of plaques scored (Lundberg, *et al.*, 1991).

The *Pfu* DNA polymerase was initially isolated directly from *P. furiosus* (Lundberg, *et al.*, 1991), but it is difficult to grow this thermophilic, anaerobic bacterium to obtain large quantities of protein. A major advance was expression of recombinant *Pfu* DNA polymerase in a baculovirus-mediated system (Mroczkowski, *et al.*, 1994). This system makes possible production of commercial amounts, but is

not as cheap and easy as an *E. coli* bacterial expression system. The *Pfu* DNA polymerase was also produced in *E. coli* expression system as a native (Lu and Erickson, 1997), or His-tagged protein (Dabrowski and Kur, 1998) giving milligram quantities of enzyme.

A *Pfu* DNA polymerase assay has been proposed by Mroczkowski (1994) for the incorporation of deoxyribonucleotides into DEAE-paper bound form, in which one unit was defined as the amount of protein that catalyzed the incorporation of 10 nmol total nucleotide into a DEAE-bound form in 30 minutes at 72 °C (Mroczkowski, *et al.*, 1994). Relative *Pfu* DNA polymerase activity, comparing the PCR results obtained from unknown activity *Pfu* DNA polymerase with that from commercial *Pfu* DNA polymerase, has also been used (Dabrowski and Kur, 1998). The activity obtained from *Pfu* DNA polymerase directly purified from *P. furiosus* was 31,713 units/mg protein (Lundberg, *et al.*, 1991), from baculovirus expression was 26,000 units/mg protein (Mroczkowski, *et al.*, 1994), from the bacterial pET11 expression system, expressed as native protein, was 22,500 units/mg protein (Lu and Erickson, 1997), and from the bacterial pET30 expression system, as His₆-tagged form, was 31,000 units/mg protein (Dabrowski and Kur, 1998).

The *Pfu* DNA polymerase protein was previously crystallized from 0.08 M ammonium sulfate, 0.05 M Na-cacodylate, pH 6.5, 0.15%(v/v) NP40, 0.05%(v/v) Tween 20 and 4.5%(w/v) polyethylene glycol 6000 by the vapor-diffusion method, yielding long rod-like single crystals. The crystals diffracted only beyond 4 Å on a synchrotron radiation source (Goldman, *et al.*, 1998).

2.2 Crystal structure of archaeal family B DNA polymerases

The first crystal structure of a family B DNA polymerase obtained was that of bacteriophage RB69 DNA polymerase (RB69 DNA polymerase) which was solved by Wang (1994). The editing complex of RB69 DNA polymerase has been reported (Shamoo and Steiz, 1999). The first crystal structure of archaeal DNA polymerase was DNA polymerase from *Thermococcus gorgonarius* (*Tgo* DNA polymerase) (Hopfner, *et al.*, 1999). Three further crystal structures of archaeal family B DNA polymerases have also been reported: Tok DNA polymerase from *Desulfurococcus* sp. Tok (Zhao, *et al.*, 1999), 9° N-7 DNA polymerase from *Thermococcus* sp. 9° N-7 (Rodriguez, *et al.*, 2000), and KOD DNA polymerase from *Pyrococcus kodakaraensis* KOD1 (Hashimoto, *et al.*, 2001). The overall structural comparisons of the four archaeal DNA polymerases are shown in figure 2.1. The structural architectures of the proteins are similar and composed of domains and subdomains, which are the N-terminal, exonuclease, polymerase domain, including the palm and fingers subdomains, and the thumb domains.

Two disulfide bonds exist in the connection site between the Palm and Fingers subdomains in all four crystal structures. The sequence alignment shown in Figure 2.2 (Hashimoto, *et al.*, 2001) suggested that DNA polymerase from *P. furiosus* (*Pfu* DNA polymerase) with an optimum growth temperature of 95 °C (similar to that of *P. kodakaraensis* KOD1 (*Pko*)) is expected to have 2 disulfide bonds as well. It also suggested that the numbers of disulfide bonds are correlated with optimum growth temperatures of organisms. DNA polymerases from *Thermococcus litoralis*, *Methanococcus jannaschii* and *Archaeoglobus fulgidus*, with optimum growth temperatures of 85, 85, and 83 °C, respectively, are expected to have one disulfide

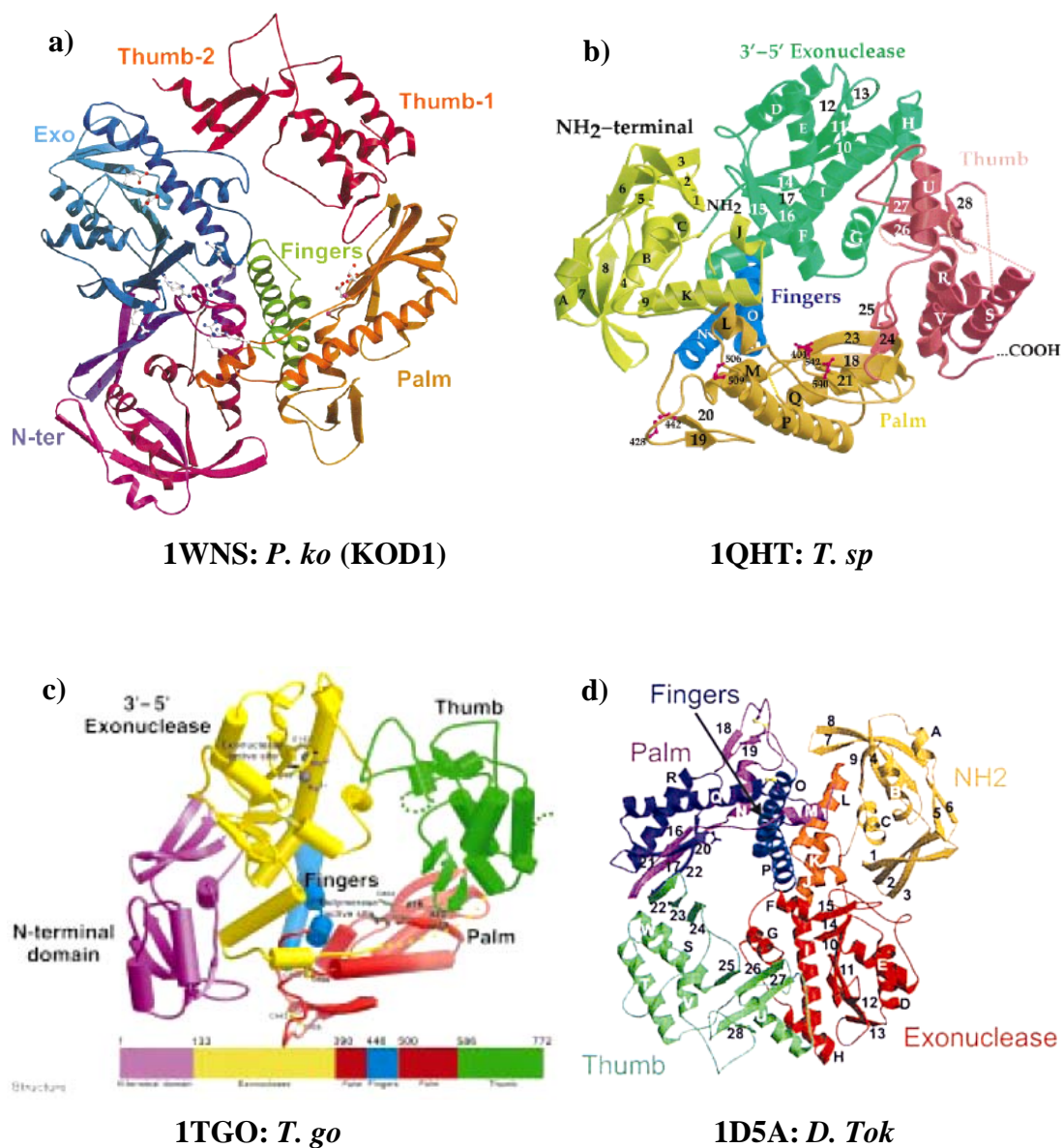


Figure 2.1 Structures of archaeal DNA polymerases. **a)** *Pyrococcus kodakaraensis* KOD1 [PDB Code: 1WNS] (Hashimoto, *et al.*, 2001). **b)** *Thermococcus sp.* 9° N-7 [PDB Code: 1QHT] (Rodriguez, *et al.*, 2000). **c)** *Thermococcus gorgonarius* [PDB Code: 1TGO] (Hopfner, *et al.*, 1999). **d)** *Desulfurococcus* strain Tok. [PDB Code: 1D5A] (Zhao, *et al.*, 1999).

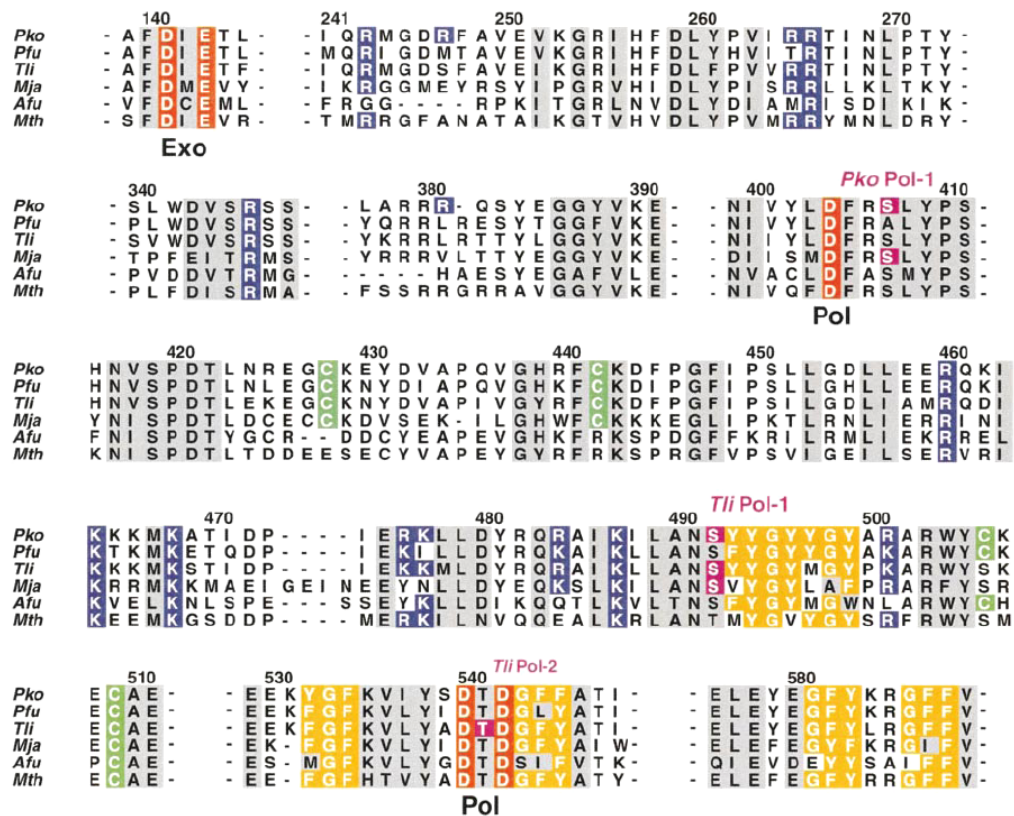


Figure 2.2 Sequence alignment of archaeal DNA polymerases. The abbreviations used as follows: *Pko*, *Pyrococcus kodakaraensis*; *Pfu*, *Pyrococcus furiosus*; *Tli*, *Thermococcus litoralis*; *Mja*, *Methanococcus jannaschii*; *Afu*, *Archaeoglobus fulgidus*; and *Mth*, *Methanobacterium thermoautotrophicum*. Homologous residues are masked in gray. Remarkable residues are highlighted in reverse type. Conserved carboxylate residues in the Exonuclease and Polymerase active sites are shown in red. Basic residues gathering in the forked-point and Fingers subdomain are shown in blue. Cysteine residues forming (or possibly forming) disulfide bonds are shown in green (Hashimoto, *et al.*, 2001).

bond, because *Pko* Cys506 is replaced by serine in *T. litoralis* and *M. Jannaschii*, and *Pko* Cys442 is replaced by arginine in *A. fulgidus*. DNA polymerase from *Methanobacterium thermoautotrophicum*, with an optimum growth temperature of 65 °C, is expected to have no disulfide bond, because *Pko* Cys428, Cys442 and Cys506 are replaced by glutamic acid, arginine and serine, respectively.

2.3 Protein Crystallization

The crystallization is the first and in many cases the trickiest part of a structure determination. Whatever is done in the following steps, the quality of the crystal sets a limit on the quality of the final models. Most important is the maximal resolution to which the crystal diffracts.

2.3.1 Principles of protein crystallization

In order to initiate crystallization, the protein solution has to be brought to a thermodynamically unstable state of supersaturation. The solution can be brought back to the stable equilibrium state through precipitation of the protein, which is the most frequent process, or through crystallization. The supersaturated state can be achieved by several techniques: evaporation of solvent molecules, change of ionic strength, change of pH, change of temperature or change of some other parameter (Bergfors, 1999).

Figure 2.3 show the phases of crystal growth, which include nucleation, growth and cessation of growth. Nucleation requires a greater protein concentration than growth. The solution needs to be supersaturated. Supersaturation might, however, lead to formation of too many nucleation points. Seeding is a way of limiting the number of nucleation points. Growth rate and crystal size depend on the

degree of super-saturation. Cessation of growth may be caused by depletion of a particular component, which actually is building up the crystal, growth defects or the flow of molecules around the crystal (Rhodes, 2006). The crystal-packing interactions involve salt bridges, hydrogen bonds, van der Waals-, dipole-dipole-, and stacking interactions.

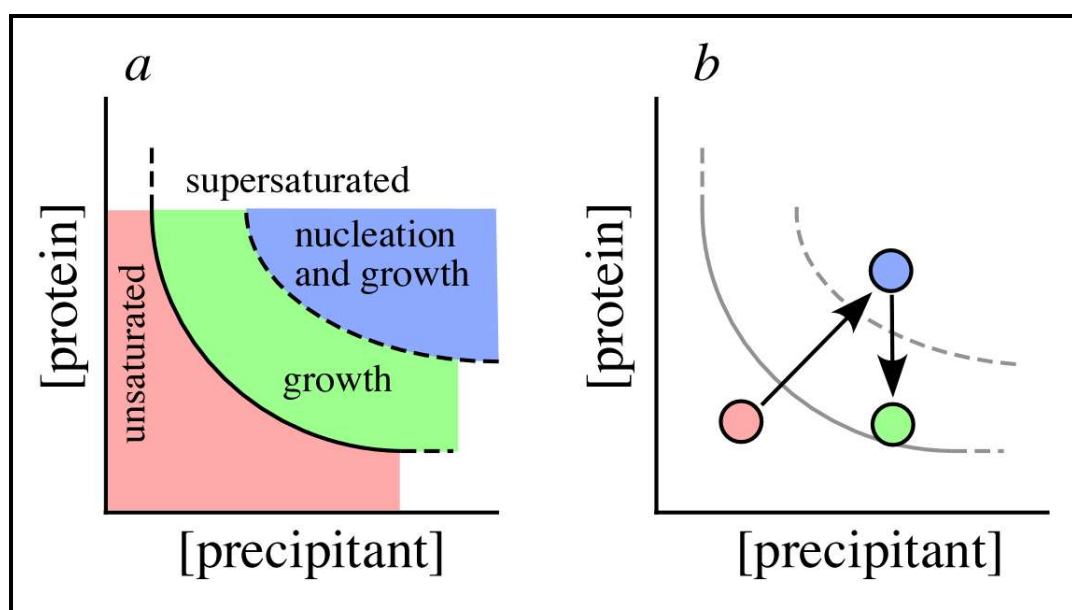


Figure 2.3 Phase diagram of crystal growth. (Rhodes, G. 2006. Resources for Readers Crystallography Made Crystal Clear: a Guide for Users of Macromolecular Models [online]. Available: <http://www.usm.maine.edu/~rhodes/CMCC/>)

2.3.2 Protein crystallization methods

Bergfors (1999) has described the methods widely used in protein crystallization including vapor diffusion, dialysis, batch, and microbatch techniques.

2.3.2.1 Vapor diffusion

The most frequently used crystallization method is the vapor diffusion technique. In this technique a small droplet of typically 2-10 μ l of the protein is

mixed with an equal or similar volume of the crystallizing solution (usually buffer, salt, and precipitant) and placed in a reservoir containing 500 – 1,000 μl of the crystallizing solution in a closed system. The difference in concentration between the drop and the reservoir drives the system toward equilibrium by diffusion, usually of water molecules, through the vapor phase. The protein becomes supersaturated and crystals start to form when drop and reservoir are at or close to equilibrium.

Vapor diffusion method can be prepared in several ways including hanging drops, sitting drops, sandwich drops, reverse vapor diffusion, and pH gradient vapor diffusion

2.3.2.2 Dialysis

Dialysis offers a way of manipulating the ionic strength that is not possible with vapor diffusion. The dialysis can be set up in various ways, e.g., in dialysis bags or collodion thimbles (also called ultra thimbles), Zeppezauer cells, or buttons. In the button the protein is placed inside the button which then covered by a dialysis membrane. Dialysis membranes are semi-permeable: They allow small molecular weight substances to diffuse in while preventing the protein from diffusing out.

Dialysis is the most effective technique for crystallization by decreasing the ionic strength. If a protein is less soluble at low ionic strength, it is often possible to crystallize it simply by dialyzing it against a weak buffer or even water. In this case, no precipitant is necessary. Unlike vapor diffusion, the protein concentration remains constant during the dialysis experiment.

2.3.2.3 Batch technique

In the batch technique, the precipitant and the target molecule solution are simply mixed. Supersaturation is achieved by mixing rather than by diffusion. It can

be also be used as a rapid micro-scale screening under microscope for the purpose of determining the solubility conditions.

2.3.2.4 Microbatch technique

Batch technique can also be performed under oil. The oil prevents evaporation and extremely small drops ($< 2 \mu\text{l}$) can be made, hence it is termed “microbatch”.

2.3.3 Symmetry in protein and macromolecular crystals

The macromolecules in biological systems very often have some type of symmetry. Using multiple copies of a single biological unit, evolution has created molecules with properties that are not easily obtained with a single unit. The simplest cases are those where two copies of a protein form a dimer, but there are cases (mainly viruses) where even thousands of identical molecules form a regular arrangement.

Proteins are asymmetric molecules formed by L-amino acids. This means that the symmetry operations must involve only rotations and translations. A rotation axis rotates an object by a fixed amount around an axis. If no translation is involved, the rotation is called an n -fold rotation axis when it rotates the object $360^\circ/n$ around the axis. The number n must be an integer, but might have any value larger than 1. When the object is both rotated and translated, we will have a helix. Well-known examples of proteins forming helical structures are actin, tubulin and the coat protein of tobacco mosaic virus.

When $n = 2$, we have two molecules forming a dimer. In dimers, a contact between molecule 1 and 2 will be repeated by the same interaction between molecule 2 and 1. In the cases where $n = 3$ or higher, there will be two different areas

of contact. These areas on the protein surface must be in the correct geometrical position to allow the interactions to occur with n-fold symmetry. In the evolution of biomolecules, the development of dimeric contacts is much more common than other types of symmetry. For example, tetramers with four-fold rotational symmetry are rare but an arrangement of four molecules with perpendicular two-fold axes (222 symmetry) is very common.

Table 2.1 Types of symmetry found in proteins and macromolecules.

Type of symmetry	Examples (PDB entry)
2	Alcohol dehydrogenase, cro repressor (5cro), Immunoglobulin G
3	Porin (1A0S), Influenza virus hemagglutinin
4	Influenza virus neuraminidase, K ⁺ channel (1K4C)
5	Serum amyloid P component
6	C-Phycocyanin
8	Light-harvesting complex 2
9	Light-harvesting complex 2 (1IJD)
11	trp RNA-binding attenuation protein
16	Light-harvesting complex 1
222	Lactate dehydrogenase, prealbumin (1BMZ)
32	Aspartate carbamoyltransferase (1D09)
422	Hemerythrin (1I4Y), Glycolate oxidase
622	Glutamine synthetase (1F1H)
72	GroEL (1DER)
23 (Tetrahedral)	Protocatechuate 3,4-dioxygenase (3PCA)
432 (Octahedral)	Ferritin (1BFR)
532 (Icosahedral)	Icosahedral viruses (STNV: 2STV), Riboflavin synthetase

Some of the symmetric arrangements create cylinders where a central hole is important for the function. In icosahedral viruses, a large number of subunits form a protein shell around the viral genome. Further information concerning the symmetry

of protein crystal forms is described in APPENDIX F.

2.4 Determination of the crystal structure of proteins

This topic is aimed to overview the current strategies and methods in determination of crystal structures of macromolecules, so called “X-ray crystallography”. The diffraction theory underlying the structure determination of a macromolecule has been well defined for a number of years and detailed texts such as Blundell and Johnson (1976), Giacovazzo (1992) and more recently Drenth (1994) are widely available. The works of Glusker and Trueblood (1972), Hammond (1997), Blow (2002), and Rhodes (2006) offer a more descriptive explanation of the subject for the non-specialist. A brief statement of the appropriate theory pertinent to this work is included where necessary and further useful equations can be found in APPENDICES E and F.

2.4.1 Data collection

In the data collection, the locations and intensities of spots in the diffraction pattern from the crystal are measured (figure 2.4). Since the diffraction is three-dimensional, the crystal is in some way rotated, and the scattered reflection intensities is detected on some sort of two-dimensional device. This raw data is then processed with computer programs that with little or no manual intervention are able to interpret the pattern as reflections and evaluate the intensity of the individual reflections.

2.4.1.1 Scattering from crystals

Diffraction refers to the effects observed when light is scattered into directions other than the original direction of the light, without change of wavelength. An X-ray photon may interact with an electron and set the electron oscillating with the X-ray

frequency. The oscillating electron may radiate an X-ray photon of the same wavelength, in a random direction, when it returns to its unexcited state. Other processes may also occur, akin to fluorescence, which emit X-rays of longer wavelengths, but these processes do not give diffraction effects.

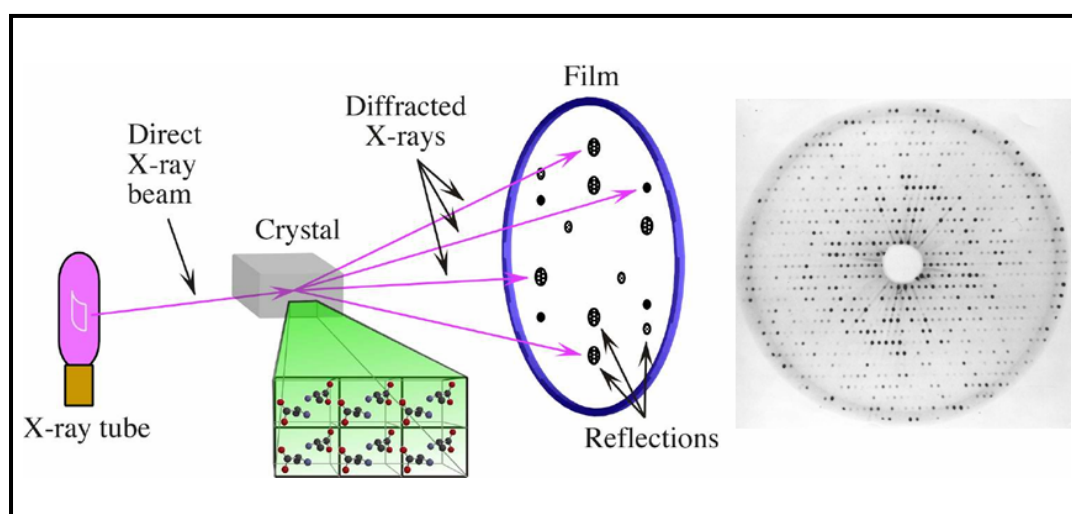


Figure 2.4 Crystallographic data collection. (Rhodes, G. 2006. Resources for Readers Crystallography made crystal clear: a guide for users of macromolecular models [online]. Available: <http://www.usm.maine.edu/~rhodes/CMCC/>)

Just as we see a red card because red light is scattered off the card into our eyes, objects are observed with X-rays because an illuminating X-ray beam is scattered into the X-ray detector. Our eye can analyze details of the card because its lens forms an image on the retina. Since no X-ray lens is available, the scattered X-ray beam cannot be converted directly into an image. Indirect computational procedures have to be used instead.

X-rays are penetrating radiation, and can be scattered from electrons

throughout the whole scattering object, while light only shows the external shape of an opaque object like a red card. This allows X-rays to provide a truly three-dimensional image. When X-rays pass near an atom, only a tiny fraction of them is scattered: most of the X-rays pass further into the object, and usually most of them come straight out the other side of the whole object. In forming an image, these “straight through” X-rays tell us nothing about the structure, and they are usually captured by a beam stop and ignored.

In diffraction by a repetitive object, rays are scattered in many directions. Each unit of the lattice scatter, but a diffracted beam arises only if the scattered rays from each unit are all in phase. Otherwise the scattering from one unit is cancelled out by another.

In two dimensions, there is always a direction where the scattered rays are in phase for any order of diffraction. In three dimensions, it is only possible for all the point of a lattice to scatter in phase if the crystal is correctly oriented in the incident beam.

The positions, amplitudes and phases of all the scattered beams from a three-dimensional crystal still provide the Fourier transform of the three-dimensional structure. But, when a crystal is at a particular angular orientation to the X-ray beam, the scattering of a monochromatic beam provides only a tiny sample of the total Fourier transform of its structure.

2.4.2 Phasing

Although the diffracted X-ray reflections from the crystal have a (relative) phase and an amplitude, our detectors are unable to record the phase difference. We want to calculate an electron density map based on our diffraction, but to be able to do

that we need to know the phase. This phase problem has to be solved in some way, and crystallographers have developed a number of methods for this. In the case of macromolecules, three methods are mainly used.

The first method used was “*isomorphous replacement*”, where heavy atoms (atoms with many electrons) are in various ways added to the molecules in the crystal. Their contribution to the scattering will be significant, even if they are (ideally) very few, compared to the great number of light atoms in the macromolecules. The position of these heavy atoms can be found, and phases for the reflections from the macromolecules can be calculated using observed amplitudes of the normal native crystal and the heavy atom derivative. The certain phase improvement methods can sometimes succeed with such phases from only one derivative, in which case the structure is said to be solved by the method of “*single isomorphous replacement*” (SIR). In practice, it commonly takes two or sometimes three or more heavy-atom derivatives to produce enough phase estimates to make the needed initial dent in the phase problem. Obtaining phase with two or more derivatives is called the method of “*multiple isomorphous replacement*” (MIR).

The second method is “*anomalous scattering*” or “*anomalous dispersion*”, a second means of obtaining phases from heavy-atom derivatives. The method takes advantages of the heavy atom’s capacity to absorb X-rays of specified wavelength. As a result of this absorption, Friedel’s law does not hold, and the symmetry-related reflections are not equal in intensity. The effect is especially large when the wavelength is close to an *absorption edge*. Absorption edges for the light atoms in the unit cell are not near the wavelength of X-rays used in crystallography, so carbon, nitrogen, and oxygen do not contribute to anomalous scattering. However, absorption

edges of heavy atoms are in this range, and if X-rays of various wavelengths are available, as often the case at synchrotron sources, X-ray data can be collected under conditions that maximize anomalous scattering by the heavy atom. This method is so called “*multiwavelength anomalous dispersion*” (MAD). By this method, all phase information can be obtained from a single crystal. A modern development of this method is the use of modified proteins where methionine residues are replaced by selenomethionines. This has opened the door to the MAD phasing for nonmetalloproteins. Moreover, several elements have a significant anomalous signal at the in house X-ray source that can be used to generate phase information from carefully measured single wavelength anomalous scattering data. This method is called “*single-wavelength anomalous dispersion*” (SAD).

The third method for phasing is “*molecular replacement*” (MR). This method entails calculating initial phases by placing a model of the known protein in the unit of the unknown protein. In turn, the method requires that the structure of the molecules studied (or part of it) is known, at least approximately. The known protein in this case is referred to as a search model. For proteins, this approximate known molecular structure is generally derived from the known structure of a related protein. If the position of the “known” molecule in the crystal can be found, approximate phases can be calculated and an electron density map for the unknown molecule can be calculated. With an increasing number of known structures, this method is now often applicable.

2.4.3 Model building

When phases have been determined, an electron density can be calculated from the measured amplitudes. An atomic model of the macromolecule is then built

in this electron density. In practice, the electron density map is often first improved by some density modification technique. The modeling originally used metal pieces in a scale of 1 inch per Å and a very large device where the electron density drawn on plastic sheets were placed behind a half-silvered mirror. The model was built in front of the mirror where it could be viewed in the density. Now modeling is performed on a graphics display and uses programs that include many software tools. These tools make it possible to efficiently build a model that fits both the electron density map and the knowledge of allowed conformations from earlier structure determination of macromolecules and small molecules. There are also procedures for more automatic model building.

The atomic model consists of list of the coordinates of (hopefully) every non-hydrogen atom of the macromolecule. Hydrogen atoms are visible only at very high resolution. The position of many hydrogen atoms is well defined by the other atoms, but others are not, and they are therefore often neglected. In high resolution studies, the hydrogen atoms are also modeled.

2.4.4 Model refinement and validation

The initial atomic model will have errors. The main reason for this is that it is based on estimated phases, and these will normally have significant errors that might make the map (at least locally) difficult to interpret. Another reason for errors in the atomic model is small inaccuracies or even mistakes at the model building. The model coordinates are therefore refined using procedures that in various ways correct these errors. These methods try to minimize the differences between the experimental data for the crystal, for example the amplitude of the reflections, F_{obs} , and the corresponding quantities calculated from the current model, F_{calc} . In most cases, the

refinement program at the same time tries to minimize deviations from standard geometry and avoid collisions of the atoms. The refined coordinates can be used to calculate improved phases and electron density maps. These maps might allow manual rebuilding of regions of the model, where the refinement program was unable to correct the model. After cycles of alternating refinement and model refinement, inclusion of water, metal ion, and in some cases, other heteroatoms is needed to be done in order to obtain the final model.

The final step in the structure determination is the validation, where various parameters (deviations of bond lengths and angles, Ramachandran plot, correspondence between calculated and observed amplitude) of the model are compared with other models to allow a decision about the quality of the model.

2.5 Quality indicator for the protein structure

2.5.1 The R factor

The first quality indicator for the structure is the R factor. The original R factor, proposed by Alan Booth (1945), which is implied when there is no subscript to the R , is used as a measure of the fit of the diffraction calculated from the structure to the observed intensity data. The R factor compares the observed structure amplitudes $|F_{\text{obs}}|$ to those calculated from the current model $|F_{\text{calc}}|$. This R factor is the ratio

$$\frac{\text{sum of differences between observed and calculated structure amplitudes}}{\text{sum of observed structure amplitudes}}$$

Or

$$R = \frac{\sum_h ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum_h |F_{\text{obs}}|}$$

As with other R factors, some authors express it as a percentage. Thus ' $R = 20\%$ ' is the same as ' $R = 0.2$ '.

The R factor is calculated over a group of reflections h , which may be all the observed reflections, or a particular group. Frequently an R factor is calculated over small ranges or 'bins' of resolution, to give an idea of the performance of the model as resolution is increased.

The R factor falls towards zero as the observed and calculated structure amplitudes agree more closely, but the values achieved depend on the resolution and degree of order of the crystals, the quality of the diffraction data, and so on.

In early stages, the R factor may be as high as 0.5. At this stage, all atoms are usually included with full weight, the coordinates are not optimized, and individual B factors have not been assigned. Even if the coordinates were perfect, the failure to assign proper B factors would have a serious effect on the R factor. The program of refinement may be watched by monitoring the R factor, and ensuring that it continues to fall.

2.5.2 The free R factor

Refinements have often been monitored by following the fall of the R factor, but this criterion was criticized by Brunger (1997), who demonstrated the bias that is introduced by using the same data for monitoring as for refinement. The reflection procedure may make false adjustments which reduce the value of R , but which do not actually improve the model.

It is then usual to select randomly a small fraction of reflections (typically 5%) which are deleted from the data used for refinement. These reflections, and only these, are used to calculate an R factor, which is called R_{free} (Brunger, 1997) and it is

defined as

$$R_{\text{free}} = \frac{\sum_{\text{test set}} \left| |F_{\text{obs}}| - |F_{\text{calc}}| \right|}{\sum_{\text{test set}} |F_{\text{obs}}|}$$

Reduction of the R_{free} factor is an unbiased estimation of the improvement to the model, because R_{free} is calculated from reflections that the refinement procedure ‘does not know’ about. For this reason R_{free} is somewhat larger than the conventional R factor. As always, cases vary, but typically R_{free} is about 1.2 times R at the end of the refinement.

The R factor varies in any case with resolution. It is often relatively high at low resolution (say as far as 8 Å). This indicates that the methods used for modeling the solvent part of the density are imperfect. This can be corrected, but rarely has a serious effect on interpretation at the level of individual amino acids. R reduces to its lowest value at intermediate resolution, and begins to rise for the weaker reflections at higher resolution. If it rises beyond 0.35 or so, it suggests that either the refinement is failing to deal correctly with the finer detail, or else that the intensities in this outermost part of the diffraction data are insufficiently accurate to give usable information.

CHAPTER III

MATERIALS AND METHODS

3.1 Materials

3.1.1 Bacterial strains

The *E. coli* strain NovaBlue (Novagen) was used for preparation of plasmid and cloning. The *E. coli* strains Rosetta (DE3) pLysS (Novagen) and BL21-Codonplus (DE3)-RIL (Stratagene) were applied to expression of His₈-tagged *Pfu* DNA polymerase.

3.1.2 Expression plasmid vector

The pSY5 plasmid (a modified pET21d) used for the construction of the expression system was kindly provided by Bob Robinson's Laboratory (BR Lab), Institute of Molecular and Cell Biology (IMCB), Singapore.

3.1.3 Enzymes

The thermostable *Pfu* DNA polymerase used for PCR amplification of *Pfu* polymerase gene from the genomic DNA template was purchased from Promega. The *Sfi*I and *Kpn*I restriction enzyme purchased from Fermentas were used for digestion of plasmid vector and PCR product in cloning. The ligation reaction of the digested plasmid vector and PCR product was performed by T4-DNA ligase enzyme purchased from Promega. The chicken egg white lysozyme was purchased from Merck.

3.1.4 Reagents and kits

The genomic DNA of *Pyrococcus furiosus* (DSM 3638) was purchased from American Type Culture Collection (ATCC, catalog number 43587D) and was used as the template for PCR amplification of DNA polymerase gene. The PCR purification kit and Plasmid preparation kit were purchased from QIAGEN. The λ DNA used as the template in the PCR efficiency assay was purchased from Fermentas.

3.2 Apparatuses

High throughput purification on AKTAXpress (GE Healthcare), including affinity, desalting, ion exchange and gel filtration chromatography allowed automated protein purification. All PCR amplifications were performed on the GeneAmp 9700 96 well (0.2ml) Thermal Cycler (Applied Biosystems). Either agarose or SDS-polyacrylamide gels were photographed with Gene-Flash Gel Doc (Syngene Bioimage). The DNA sequencing was done on the ABI Prism 373 Genetic Analyzer (Applied Biosystems). The centrifugation of lysis cell was performed by Refrigerated High Speed Centrifuge, Model RC 26 PLUS (Sorvall, Dupont). The NanoDrop ND-1000 Spectrophotometer (Biofrontier) was used for measurement of DNA and protein concentration. Crystallization screening was done by the Screenmaker 96+8 crystallization robot (Innovadyne).

3.3 Methods

3.3.1 Cloning of the DNA polymerase gene from *P. furiosus*

Based on the known DNA sequence of *Pfu* DNA polymerase gene (GenBank Accession No. D12983), the specific primers for the *Pfu pol* gene, containing *Sfi*I and

KpnI restriction recognition sites were designed and used for the amplification from genomic DNA of *P. furiosus*. The PCR reaction was carried out in a 50 µl solution that consisted of *P. furiosus* genomic DNA (200 ng), 1 µl (10 µM) of each primer (Set 1 without restriction recognition sites: PfuFL-for 5'-ATG ATT TTA GAT GTG GAT TAC ATA ACT G-3' and PfuFL-rev 5'- TAG GAT TTT TTA ATG TTA AGC CAG GAA GTT AGG-3' or Set 2 with restriction recognition sites: PfuFL-for-*SfiI* 5'-GGG CCC GGG CGG CCG ATG ATT TTA GAT GTG GAT TAC ATA ACT G-3' and PfuFL-rev-*KpnI* 5'-GGG GTA CCC CCT AGG ATT TTT TAA TGT TAA GCC AGG AAG TTA GG-3'), 1 µl (10 mM) of dNTPs, 5 µl of 10X PCR buffer (200 mM Tris-HCl, pH 8.8 at 25 °C, 100 mM KCl, 100 mM (NH₄)₂SO₄, 20 mM MgSO₄, 1 mg/ml nuclease-free BSA, 1% Triton X-100), and 3 U of thermostable *Pfu* DNA polymerase (Promega). In nested PCR, 1 µl of PCR product solution amplified with primer set 1 was used as the template for amplification with primer set 2. An initial cycle of 2 min at 95 °C, followed by 30 cycles of denaturation for 1 min at 95 °C, annealing for 30 s at 65 °C, and elongation for 2 min 30 s at 73 °C, and a final extension step at 73 °C for 7 min were performed in an automated thermal cycler (Applied Biosystems GeneAmp 9700). The amplification products were analyzed by electrophoresis on a 1% agarose gel stained with ethidium bromide.

Specific PCR products of approximately 2370-bp (2.37 kb) were obtained and cloned into pSY5, a modified pET21d plasmid expression vector at *SfiI* and *KpnI* restriction sites. First, the PCR product was purified by PCR purification kit (QAIGEN). Both the pSY5 plasmid vector and PCR product were digested by *SfiI* and *KpnI* restriction enzymes (Fermentas) according to the company's manual. The ligation was performed in 20 µl reaction. One hundred nanograms of the digested

vector was incubated together with 174 ng of PCR product in solution containing 2 μ l of 10X T4-DNA ligase buffer (300 mM Tris-HCl, pH 7.8 at 25°C, 100 mM MgCl₂, 100 mM DTT, and 10 mM ATP) and 3 U of T4-DNA ligase enzyme (Promega) at room temperature overnight. The recombinant *Pfu* plasmid was initially propagated in the *E. coli* NovaBlue (Novagen) and prepared by plasmid preparation kit (QIAGEN) according to the company's manual. The identity of the cloned *Pfu* DNA polymerase was confirmed by DNA sequencing, using the BigDye Terminator kit on an ABI Prism 373 Genetic Analyzer.

3.3.2 Expression of recombinant His₈-tagged *Pfu* DNA polymerase

Small-scale expression of the *Pfu* polymerase from the plasmid were initially compared in *E. coli* strains Rosetta (DE3) pLysS (Novagen) and BL21-Codonplus (DE3)-RIL (Stratagene). Cells were grown at 37 °C in 10 ml LB containing 100 μ g/ml ampicillin and 34 μ g/ml chloramphenicol to an A₆₀₀ of 0.6. IPTG was then added to a final concentration of 1 mM to induce the recombinant protein expression under the control of the T7 promoter for 5 h at 37 °C and 18 h at 25 °C. The protein expression in *E. coli* strain Rosetta (DE3) pLysS (Novagen) was scaled up and grown in 2-litre shake-flasks according to the procedure explained previously. The induction was carried out at 25 °C for 18 h.

3.3.3 Purification of recombinant His₈-tagged *Pfu* DNA polymerase

The cell was harvested by centrifugation at 4,500 rpm for 30 min. The pellet was resuspended in 100 ml of Ni²⁺-Binding buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, and 20 mM Imidazole). The cells were disrupted by sonication following lysozyme treatment with final concentration of 1 mg/ml for 1 h at 4 °C, the cell debris was removed by centrifugation at 18,000 rpm for 45 min, and the supernatant was

collected. The supernatant was then subjected to the automated AKTAXpress Purification System (GE Healthcare), including affinity (Ni^{2+} HisTrap_FF_1ml_Xpress column), desalting (HiPrep_26/10_desalting column), anion exchange (Resource_Q_1ml column) and gel filtration (HiLoad_16/60_Superdex200_prep grade column) chromatography for protein purification. The supernatant containing the soluble protein in Ni^{2+} -Binding buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, and 20 mM Imidazole) was loaded into the Ni^{2+} column and washed with 40 ml of Ni^{2+} -Binding buffer. The protein was then eluted with 5 ml of Ni^{2+} -Elution buffer (50 mM Tris-HCl, pH 8.8, 500 mM NaCl, and 500 mM Imidazole) and transferred with Anion Exchange Binding buffer (50mM Tris-HCl, pH 8.8) to the desalting column in order to remove NaCl from the solution. Then, the NaCl-free protein solution was loaded into the anion exchange column, washed with Anion Exchange Binding buffer and eluted with Anion Exchange Elution buffer (50 mM Tris-HCl, pH 8.8, 1 M NaCl) up to about 60% gradient or so that no protein was detected. After that, the protein solution was transferred to the gel filtration column, the final step in protein purification, with Gel Filtration buffer containing 50 mM Tris-HCl, pH 8.5 and 150 mM NaCl. The purity of the protein was analyzed by SDS-PAGE. Five microlitres of the protein fractions were mixed with 5 μl of 5X loading buffer (225 mM Tris-HCl, pH 6.8, 50% glycerol, 5% SDS, 0.05% bromophenol blue, 250 mM DTT) and run on 15% SDS-polyacrylamide gel at 190 volts for 50 min. The gel was stained with staining buffer (0.025% Coomassie Brilliant Blue R250, 40% (v/v) methanol, and 7% (v/v) acetate acid) for 10 min, destained in destaining buffer (40% (v/v) methanol and 7% (v/v) acetate Acid) overnight. The gel was photographed with Gene-Flash Gel Doc (Syngene Bioimage). The concentration of

purified His₈-tagged *Pfu* DNA polymerase was determined from ultraviolet absorption on the NanoDrop ND-1000 Spectrophotometer (Biofrontier), using the extinction coefficient for $A_{280} = 0.74$ for 1 mg/ml, calculated from the number of Trp and Tyr residues in the sequence (using the Protean program of DNA star, Madison, WI). The His₈-tagged *Pfu* DNA polymerase was then concentrated to 15 and 20 mg/ml by protein concentrating column Vivaspin (Vivascience).

3.3.4 Relative DNA polymerase activity assay

The purified His₈-tagged *Pfu* DNA polymerase was diluted to 0.4 µg/µl in *Pfu* storage buffer (50 mM Tris-HCl, pH 8.2, 0.1 mM EDTA, 1 mM DTT, 0.1% NP-40, 0.1% Tween 20, 50% (w/v) glycerol). The relative DNA polymerase activities were determined by comparing band intensities of the PCR products obtained with His₈-tagged *Pfu* DNA polymerase and the commercial *Pfu* DNA polymerase (Promega). The pVLN1 plasmid-containing the Plant Villin1 (*VLN1*) gene was used as the DNA template in 50 µl PCR reactions for both the His₈-tagged *Pfu* DNA polymerase and the commercial *Pfu* DNA polymerase (Promega). The primers (For_VLN1 5'-GGG CCC GGG CGG CCG ATG TCT AGG CTA AGT AAA GAC ATT GAT TC-3'; Rev_VLN1M 5'-GGG GTA CCC CTT ATC CTT TTT GTT TGA ACA ACG CAG CCA C-3') were applied to amplify the 1,150-bp fragment of *VLN1* gene. One microlitre of plasmid DNA (10 pg) was combined with 5 µl of 10X reaction buffer (200 mM Tris-HCl pH 8.8 at 25 °C, 100 mM KCl, 100 mM (NH₄)₂SO₄, 20 mM MgSO₄, 1 mg/ml nuclease-free BSA, 1% Triton X-100), 1 µl (10 µM) of each primer, and 1 µl (10 mM) of dNTPs. This mixture was supplemented with different amount of His₈-tagged *Pfu* DNA polymerase (0.2, 0.4, 0.8, 1.2, and 1.6 µg) or 3 U of commercial *Pfu* polymerase (Promega). The amplification reactions were performed

with an automated thermal cycler (Applied Biosystems GeneAmp 9700). The protocol was 2 min at 95 °C; 30 cycles of 1 min at 95 °C, 30 s at 65 °C, and 2 min 30 s at 73 °C; followed by 73 °C for 7 min.

Five microlitres of PCR products were separated electrophoretically on a 1% agarose gel stained with ethidium bromide. The DNA was visualized on an ultraviolet transilluminator and photographed with Gene-Flash Gel Doc (Syngene Bioimage). The band intensities were determined and compared by Quantity One® Software (Bio-Rad).

3.3.5 Determination of His₈-tagged *Pfu* DNA polymerase thermostability

For the thermostability studies, the His₈-tagged *Pfu* DNA polymerase was heat-treated at 97.5 °C for 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 23 h prior to the PCR reaction. After various incubation periods the enzyme samples (5 U) were withdrawn and kept at 4 °C overnight before assaying except for the 23-h treatment which was subjected immediately to the DNA polymerase activity assay using specific PCR amplification with pVLN1 template as described earlier.

3.3.6 Examination of His₈-tagged *Pfu* DNA polymerase PCR efficiency

The method used was that described earlier by Wang, *et al.* (2004) with some modification. One hundred and fifty nanograms of λ DNA (Fermentas) was used as template to assess the relative PCR efficiency of His₈-tagged *Pfu* DNA polymerase compared with that of commercial native *Pfu* polymerase (Promega) in PCR reactions. A set of primers listed in table 3.1 was used to amplify amplicons of 0.5, 1, 2, 5, 8, 10, 12 and 15 kb in size from the template in a 50 μ l reaction. The PCR reaction solution consisted of 1 μ l of λ DNA (150 ng), 5 μ l of 10X reaction buffer (200 mM Tris-HCl, pH 8.8 at 25 °C, 100 mM KCl, 100 mM (NH₄)₂SO₄, 20 mM

MgSO₄, 1 mg/ml nuclease-free BSA, 1% Triton X-100), 1 µl (10 µM) of each primer, 1 µl (10 mM) of dNTPs and 5 U of either enzymes (His₈-tagged *Pfu* DNA polymerase or Promega native *Pfu* polymerase). The amplification reactions were performed with an automated thermocycler (Applied Biosystems GeneAmp 9700) according to the following scheme: an initial cycle of 20 s at 95 °C was followed by 20 cycles of denaturation for 5 s at 94 °C, annealing and elongation of primers for 2.5 min at 73 °C, and a final extension step at 73 °C for 7 min. Upon completion of the PCR, 5 µl of the PCR product solution was mixed with loading buffer and loaded onto 1% agarose gel stained with ethidium bromide. The gel was visualized on an ultraviolet transilluminator, and photographed.

The heat-treated His₈-tagged *Pfu* DNA polymerase at 23 h was also tested for its PCR efficiency, using the method explained above.

3.3.7 Crystallization of His₈-tagged *Pfu* DNA polymerase

His₈-tagged *Pfu* DNA polymerase solutions with concentrations of 15 and 20 mg/ml either non-heated or heated at 97.5 °C for 30 min were subjected to screening for probable crystallization conditions with commercially available protein crystallization screening kits at 15 °C. The formulations of the commercial protein crystallization screening kits are shown in APPENDIX D. The non-heated enzyme was screened with The Pegs (QIAGEN), Ozma PEG1K&4K, Ozma PEG8K&10K, and CryO I/II (Emerald Biosystems), whereas JBScreen HTS I, HTS II (JENA Bioscience); Index and High Throughput (HAMPTON Research); The Pegs (QIAGEN); Ozma PEG1K&4K, Ozma PEG8K&10K, WIZARD I/II, and CryO I/II (Emerald Biosystems) were used for the heated enzyme. The crystallization screening was done by the sitting drop, vapor diffusion technique on a ScreenMaker 96+8

crystallization screening robot. The optimizations were done accordingly to the results obtained from the screening conditions. The optimization of non-heated His₈-tagged *Pfu* DNA polymerase was carried out with the optimization solutions stated in APPENDIX D.

Table 3.1 Primers used in extension efficiency assay.

Primer name	Primer sequence
PfuExL30350F	5'-CCTGCTCTGCCGCTTCACGC-3'
PfuExL-0.5R (0.5 kb)	5'-TCCGGATAAAAACGTCGATGACATTTGC-3'
PfuExL-1R (1 kb)	5'-GATGACGCATCCTCACGATAATATCCGG-3'
PfuExL-2R (2 kb)	5'-CCATGATTCAGTGTGCCCCGTCTGG-3'
PfuExL-5R (5 kb)	5'-CGAACGTCGCGCAGAGAAACAGG-3'
PfuExL-8R (8 kb)	5'-GCCTCGTTGCGTTTGTGTTGCACG-3'
PfuExL-10R (10 kb)	5'-GCACAGAAGCTATTATGCGTCCCCAGG-3'
PfuExL-12R (12 kb)	5'-TCTTCCTCGTGCATCGAGCTATTCGG-3'
PfuExL-15R (15 kb)	5'-CTTGTTCTTTGCCGCGAGAATGG-3'

3.3.8 Data collection and processing from the His₈-tagged *Pfu* DNA polymerase crystals

X-ray diffraction measurements were performed at the beamline BL13B1 of the National Synchrotron Radiation Research Center (NSRRC), Hsinchu, Taiwan. Each crystal of *Pfu* DNA polymerase was mounted with a Cryoloop (Hampton) from a drop of mother liquid; the crystal was then rapidly transferred to soak in the

cryosolution (mother liquor with 20% glycerol) for 2-3 min and flash-cooled in liquid N₂ before exposing to the 100 K N₂ gas stream. The data collection processes were controlled by Blu-Ice 4 software. The diffraction data were detected by the ADSC Quantum-315 CCD detector. Data to a maximum resolution of 3.0 Å was collected at a crystal to detector distance of 450 mm (applicable for collecting data to 3.0 Å resolution) using 1° oscillations with 60 sec dose equivalent exposure times for 180°.

X-ray diffraction pattern were visualized by program XDISPLAYF. The bulk of diffraction data were then integrated and corrected for Lorentz and polarization effects with the program DENZO (Otwinowski, 1993). Scaling and merging of data were achieved using program SCALEPACK (Otwinowski, 1993). These programs were implemented in the HKL2000 program package (Otwinowski and Minor, 1997).

3.3.9 Preliminary structure determination of the His₈-tagged *Pfu* DNA polymerase

A number of programs which are available from the CCP4 program package (CCP4, 1979) were used for the determination of preliminary crystal structure.

3.3.9.1 Phasing

The method used for phasing was “Molecular Replacement” with the known structure of DNA polymerase from *Pyrococcus kodakaraensis* KOD1 (PDB Code: 1WNS) previously solved by Hashimoto (2001) as the search model. This protein has 79% amino acid sequence identity to the *Pfu* DNA polymerase. The phasing was carried out using the MOLREP (Vagin and Teplyakov, 1997) and AMORE (Navaza, 1994) program. Data in the resolution range of 20.0 – 3.0 Å were used in both the rotation and translation functions.

3.3.9.2 Model building and refinement

The resulting model was subjected to refinement using program REFMAC5 (Murshudov, 1996). Electron density map was displayed and model coordinates fitted using the interactive computer graphics programs O (Jones, *et al.*, 1991) and COOT (Emsley and Cowtan, 2004). The model was rebuilt, with insertion of the correct amino acids. The PYMOL program was used in preparing the figures.

CHAPTER IV

RESULTS AND DISCUSSION

4.1 Cloning of the DNA polymerase gene from *P. furiosus*

Using 2 sets of primers, without (set 1) and with (set 2) restriction recognition sites, to amplify *Pfu* DNA polymerase gene from genomic DNA of *P. furiosus*, the PCR products of approximately 2.37 kb could only be obtained with the set 1-primers. This might be due to the long uncomplimentary region of the primer set 2 (Figure 4.1a). One microlitre of the PCR product from the reaction with primer set 1 was then used as the template for amplification with primer set 2 in nested PCR. The result showed that the 2.37 kb PCR products could be obtained (Figure 4.1b).

The PCR products of approximately 2.37 kb in size were successfully cloned into the pSY5 plasmid vector, which was designed to provide a His₈-tag to the N-terminus of the *Pfu* DNA polymerase sequence. DNA sequencing of the recombinant pSY5 plasmid showed 100% identity to the known DNA sequence of *Pfu* DNA polymerase gene (GenBank Accession No. D12983).

4.2 Expression of recombinant His₈-tagged *Pfu* DNA polymerase

The small-scale expression of the *Pfu* polymerase from the plasmid was compared in *E. coli* strains Rosetta (DE3) pLysS (Figure 4.2) and BL21-Codonplus (DE3)-RIL (Figure 4.3), induced by 1mM IPTG at 25 °C for 18 h and 37 °C for 5 h. The small-scale expression in *E. coli* Rosetta (DE3) pLysS in Figure 4.2 showed that

the induction at 25 °C, 18 h (lane 8) resulted in higher expression than that from 37 °C, 5 h (lane 10). There was no expression of protein in these uninduced samples, lanes 7 and 9, which suggested that *E. coli* strain Rosetta (DE3) pLysS had strong control of the T7 promoter. On the other hand, in the small-scale expression from *E. coli* BL21-Codonplus (DE3)-RIL, there was some expression of protein in these uninduced samples (Figure 4.3: lane 7 and 9). This suggested that *E. coli* strain BL21-Codonplus (DE3)-RIL could not control the T7 promoter as well as the *E. coli* Rosetta (DE3) pLysS. However, similarly to Rosetta (DE3) pLysS, the BL21-Codonplus (DE3)-RIL gave higher expression when induce at 25 °C for 18 h.

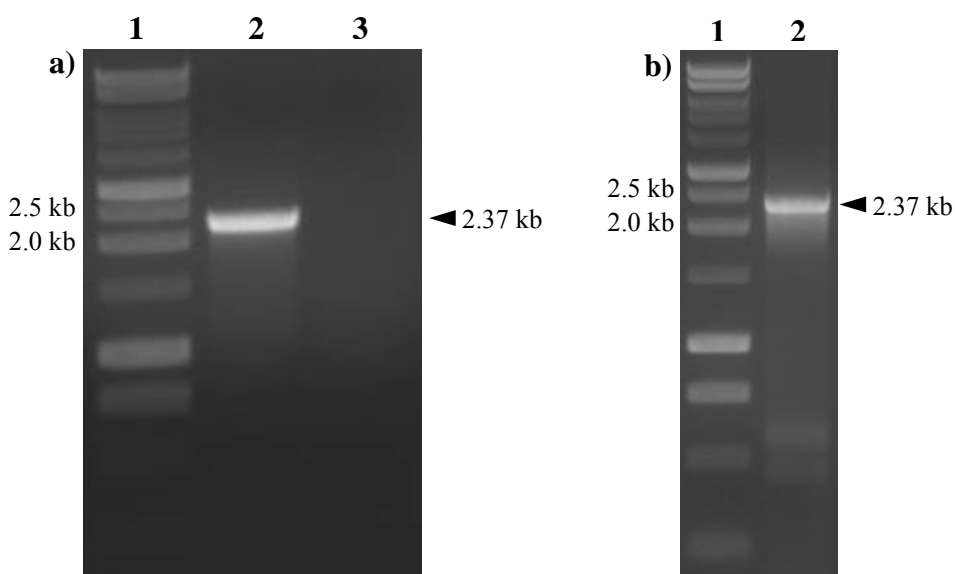


Figure 4.1 PCR products generated by the use of different primers. Genomic DNA from *P. furiosus* was used as template. Lane 1: 1 kb Marker; Lane 2: Primer set 1; Lane 3: Primer set 2 (a). Nested PCR using PCR product from lane 2 in Figure (a) as template and primer set 2 (b).

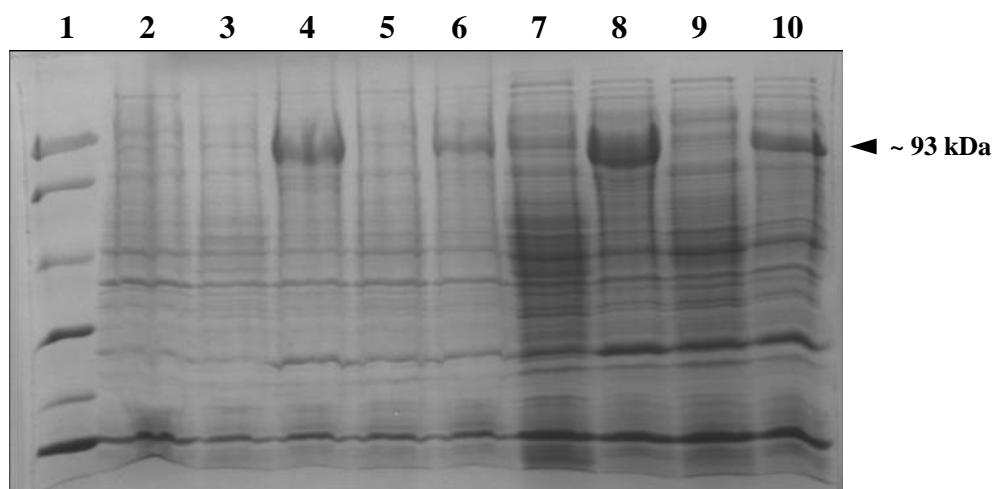


Figure 4.2 Small-scale expression of recombinant His₈-tagged *Pfu* DNA polymerase in *E. coli* strain Rosetta (DE3) pLysS analyzed by SDS-PAGE. Lane 1: Low-range protein molecular weight marker; Lane 2: Control*; Lane 3: Uninduced cell pellet** at 25 °C; Lane 4: Induced cell pellet*** at 25 °C; Lane 5: Uninduced cell pellet** at 37 °C; Lane 6: Induced cell pellet*** at 37 °C; Lane 7: Uninduced soluble⁺ at 25 °C; Lane 8: Induced soluble⁺⁺ at 25 °C; Lane 9: Uninduced soluble⁺ at 37 °C; Lane 10: Induced soluble⁺⁺ at 37 °C.

Due to the higher expression (compared lane 8 in Figures 4.2 and 4.3) and better control of the T7 promoter (compared lane 7 in Figures 4.2 and 4.3), the *E. coli* strain Rosetta (DE3) pLysS and induction at 25 °C for 18 h was chosen for scaling up.

* Whole cells which were collected when OD₆₀₀ = 0.6 without further culture.

** Whole cells which were further cultured after OD₆₀₀ = 0.6 without induction.

*** Whole cells which were induced by 1mM IPTG.

⁺ Supernatants from cells that were cultured after OD₆₀₀ = 0.6 without induction.

⁺⁺ Supernatants from cells that were induced by 1 mM IPTG.

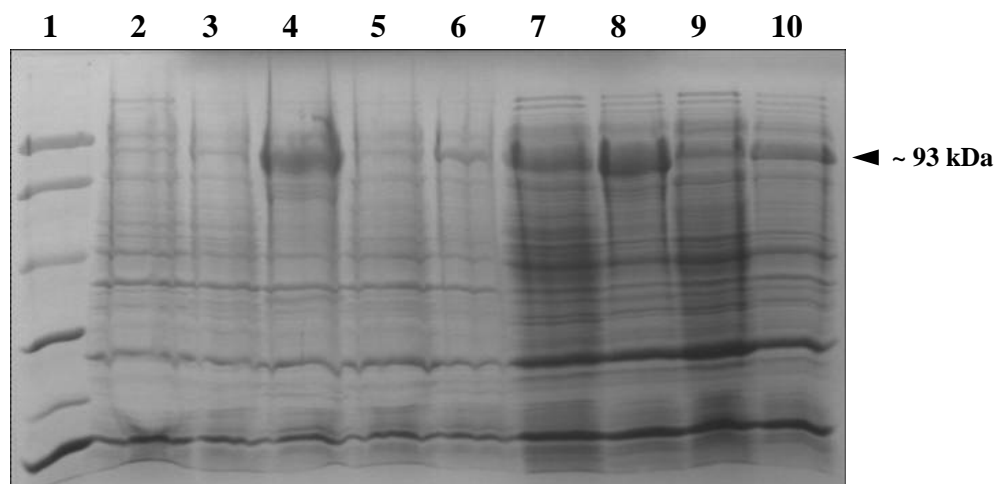


Figure 4.3 Small-scale expression of recombinant His₈-tagged *Pfu* DNA polymerase in *E. coli* strain BL21-Codonplus (DE3)-RIL analyzed by SDS-PAGE. The lanes are the same as those describe in Figure 4.2.

4.3 Purification of recombinant His₈-tagged *Pfu* DNA polymerase

The large scale (2 litres) expressed recombinant His₈-tagged *Pfu* DNA polymerase was purified with the automated AKTAXpress Purification System which had Ni²⁺ column (affinity) in the first step of purification, following by desalting, anion exchange and gel filtration chromatography. The results showed that the protein could be purified by this system which indicated that the His₈-tag at the N-terminus of the recombinant *Pfu* DNA polymerase according to the plasmid design was present for the purification. After gel filtration chromatography, the protein was separated into two peaks including a small peak, which was thought to be aggregated protein and a latter big peak of the His₈-tagged *Pfu* DNA polymerase. The fractions from only the latter big peak were pooled and used for later experiments. The purity of protein after gel filtration was at a satisfactory level (Figure 4.5).

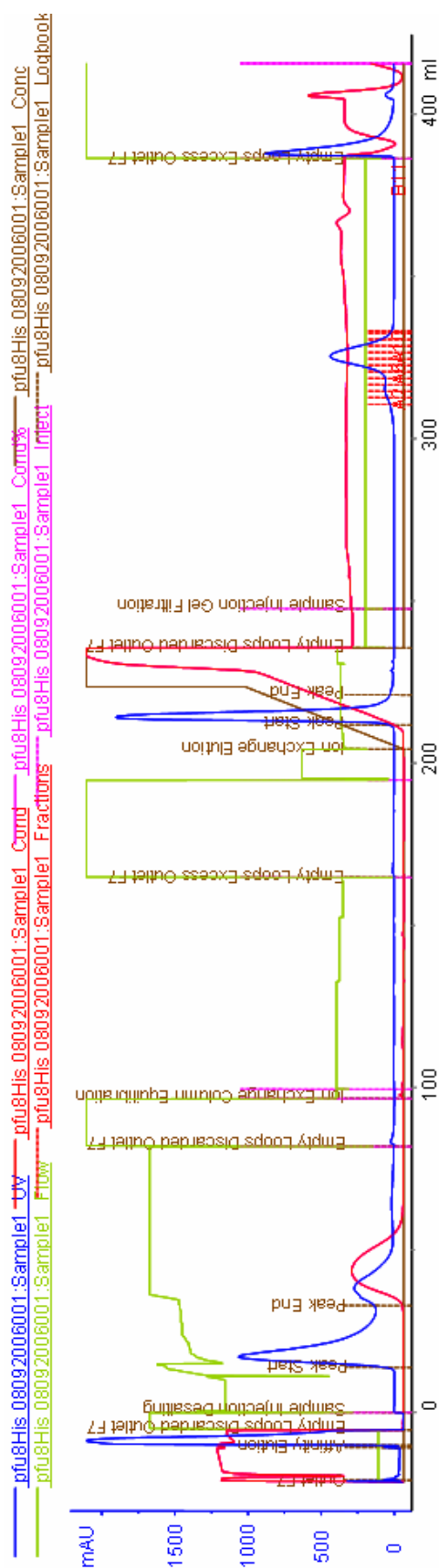


Figure 4.4 Purification profile of His₈-tagged *Pfu* DNA polymerase. The first step in purification was Ni²⁺ column (affinity), following by desalting, anionic exchange and gel filtration column.

The purified protein yield was 38 milligram protein. The purified protein was then used for activity assays and protein crystallization.

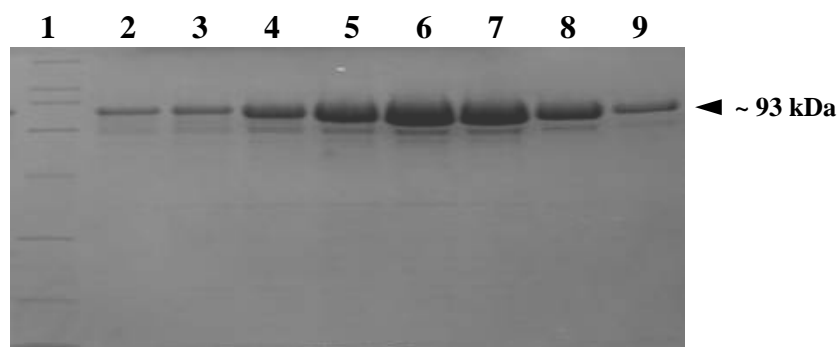


Figure 4.5 Purity of the recombinant His₈-tagged *Pfu* DNA polymerase after purification by automated AKTAXpress Purification System assessed by SDS-PAGE. Lane 1: Board-range protein molecular weight marker. Lanes 2-9: Purified protein fraction in the latter big peak.

4.4 Relative DNA polymerase activity assay

The relative DNA polymerase activity was determined by comparing the PCR results obtained from His₈-tagged *Pfu* DNA polymerase and the Promega commercial *Pfu* DNA polymerase using specific PCR amplification of the *VLNI* gene as a target. As shown in Figure 4.6 and Figure 1B in APPENDIX B, the use of 0.2 µg of His₈-tagged *Pfu* DNA polymerase (lane 2) gave rise to the band intensity of 2-fold higher than that from using 3 U of the Promega *Pfu* DNA polymerase (lane 7). This, in turn, suggested that 0.2 µg of His₈-tagged *Pfu* DNA polymerase has 6 U of enzyme or 30,000 U per mg protein, which is comparable to the previously reported activities of *Pfu* DNA polymerase directly purified from *P. furiosus* of 31,713 U/mg protein (Lundberg, *et al.*, 1991), from baculovirus expression system of 26,000 U/mg protein

(Mroczkowski, *et al.*, 1994), from the bacterial pET11 expression system, expressed as native protein, of 22,500 U/mg protein (Lu and Erickson, 1997), and from the bacterial pET30 expression system, as His₆-tagged form, of 31,000 U/mg protein (Dabrowski and Kur, 1998).

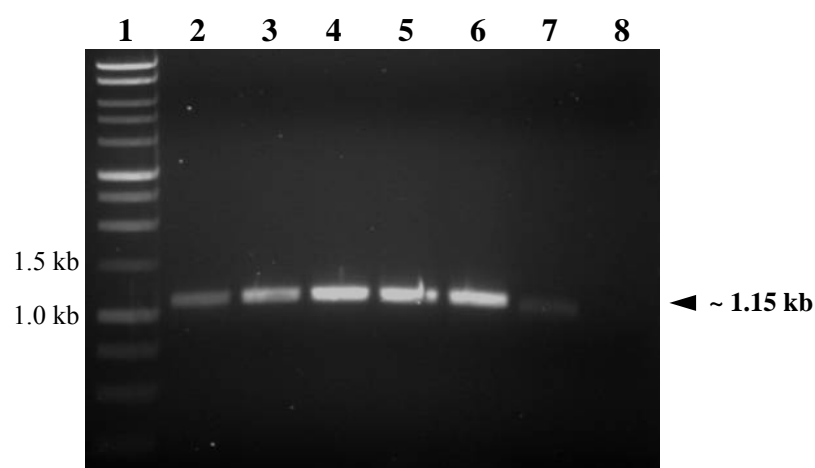


Figure 4.6 The relative DNA polymerase activity assay of His₈-tagged *Pfu* DNA polymerase after purification by the automated AKTAXpress Purification System. Lane 1: 1 kb marker; Lanes 2-6: 0.2, 0.4, 0.8, 1.2, 1.6 µg, respectively of purified His₈-tagged *Pfu* DNA polymerase; Lane 7: 3 U of Promega *Pfu* DNA polymerase; Lane 8: Control, no enzyme added in the PCR reaction.

4.5 The determination of His₈-tagged *Pfu* DNA polymerase thermostability

After heat-treating His₈-tagged *Pfu* DNA polymerase at 97.5 °C for 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 23 h, the enzyme samples were tested in PCR. The results shown in Figure 4.7 indicated that the His₈-tagged *Pfu* DNA polymerase was still very stable and active even after incubating at high temperature for at least up to 23 h,

which was the maximum duration tested.

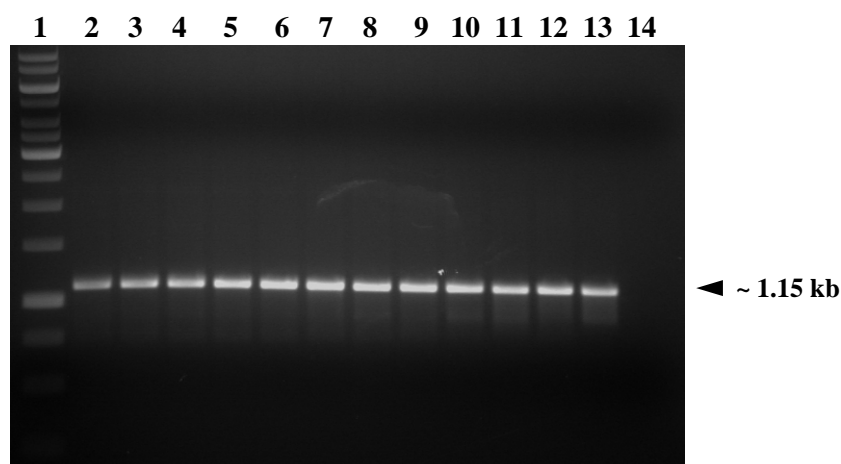


Figure 4.7 Thermostability of His₈-tagged *Pfu* DNA polymerase. The His₈-tagged *Pfu* DNA polymerase was heat-treated at 97.5 °C for 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, and 23 h (lanes 2-13, respectively) before PCR. Lane 1: 1 kb marker; Lane 14: control (no enzyme added).

4.6 The examination of His₈-tagged *Pfu* DNA polymerase PCR efficiency

The performance *in vitro* of thermostable DNA polymerases, such as in PCR or cycle sequencing applications is important. Higher performance allows the polymerase to incorporate more nucleotides per binding event, it should allow a more efficient *in vitro* replication of the template strand during each PCR cycle. Consequently, shorter extension time may be required to amplify the same target or a much longer target could be amplified with the same extension time, indicating the higher PCR efficiency (Wang, *et al.*, 2004). In this experiment, the PCR efficiency of recombinant His₈-tagged *Pfu* DNA polymerase was compared with directly extracted

commercial Promega *Pfu* DNA polymerase. Five units of each enzyme was used in amplification of various sizes of amplicons, ranging from 0.5 to 15 kb from 150 ng of λ DNA template, using an extension time of 2.5 min.

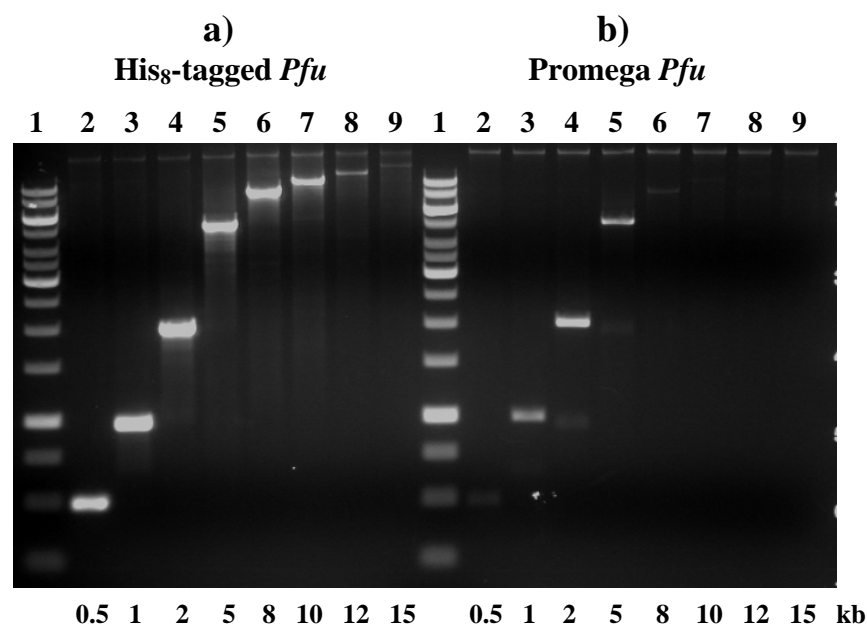


Figure 4.8 Comparison of PCR efficiency of 5 U of the recombinant His₈-tagged *Pfu* DNA polymerase (a) and the commercial Promega *Pfu* DNA polymerase (b). One hundred and fifty nanograms of λ DNA was used as the template and the sizes of the amplicons are indicated at the bottom in kb. The thermal cycling protocol was 20 s at 95 °C; 20 cycles of 5 s at 94 °C, 2.5 min at 73 °C; and 73 °C for 7 min. Lane 1: 1 kb marker; Lanes 2-9: amplicons of 0.5, 1, 2, 5, 8, 10, 12, and 15 kb in size, respectively.

The results shown in figure 4.8 indicate that at an extension time of 2.5 min the recombinant His₈-tagged *Pfu* DNA polymerase was able to amplify the PCR product as long as 15 kb whereas the commercial Promega *Pfu* DNA polymerase could amplify only up to 8 kb amplicon. In turn, the recombinant His₈-tagged *Pfu*

DNA polymerase has higher PCR efficiency than the commercial Promega *Pfu* DNA polymerase. This result suggested that the His₈-tag and some extra amino acids at the linker might be helpful in PCR efficiency of the His₈-tagged *Pfu* DNA polymerase. However, this hypothesis has not been proven.

Twenty-three hour heat-treated His₈-tagged *Pfu* DNA polymerase was also examined for its' PCR efficiency. The results shown in Figure 4.9 indicated that the heat-treated His₈-tagged *Pfu* DNA polymerase was still very stable and active for amplifying less than 8 kb PCR product. Heat-treating at 97.5 °C for 23 h might cause some damage in the enzyme that it did lose some efficiency.

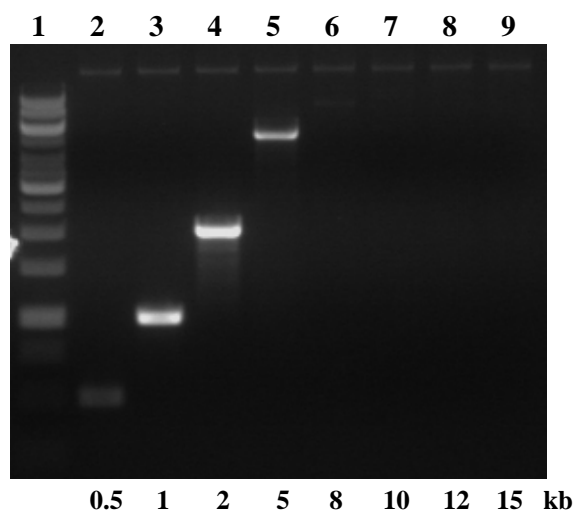


Figure 4.9 PCR efficient of 23 h heat-treated His₈-tagged *Pfu* DNA polymerase. One hundred and fifty nanograms of λ DNA was used as the template and the sizes of the amplicons are indicated at the bottom. The thermal cycling protocol and description of the lanes are those given in Figure 4.7.

Table 4.1 Hit conditions at 15 °C for crystallization of non-heated His₈-tagged *Pfu* DNA polymerase.

Screen name	Well no.	Compositions
Ozma 8K10K	A10	20%(w/v) PEG 8000, 200 mM Ammonium sulfite
	B4	20%(w/v) PEG 8000, 200 mM Lithium chloride
	B12	20%(w/v) PEG 8000, 200 mM Magnesium sulfate
	C12	20%(w/v) PEG 8000, 200 mM Sodium chloride
	D5	20%(w/v) PEG 8000, 200 mM Sodium isothiocyanate
	D9	20%(w/v) PEG 8000, 200 mM Sodium sulfate
	E9	10%(w/v) PEG 10,000, 200 mM Ammonium sulfate
	F9	10%(w/v) PEG 10,000, 200 mM Magnesium chloride
	G3	10%(w/v) PEG 10,000, 200 mM Potassium fluoride
	G7*	10%(w/v) PEG 10,000, 50 mM Potassium phosphate (monobasic)
	G8	10%(w/v) PEG 10,000, 200 mM Potassium sulfate
	H9	10%(w/v) PEG 10,000, 200 mM Sodium sulfate
Ozma 1K4K	A7*	30%(w/v) PEG 1000, 200 mM Potassium phosphate (monobasic)
	A12	30%(w/v) PEG 1000, 200 mM Calcium chloride
	C7	30%(w/v) PEG 1000, 200 mM Potassium phosphate (monobasic)
	C8	30%(w/v) PEG 1000, 200 mM Potassium sulfate
	D9	30%(w/v) PEG 1000, 200 mM Sodium sulfate
	E7	20%(w/v) PEG 4000, 200 mM Ammonium phosphate (monobasic)
	E11	20%(w/v) PEG 4000, 200 mM Calcium acetate
	H5	20%(w/v) PEG 4000, 200 mM Sodium isothiocyanate
Cryo I/II	C3	30% (v/v) PEG-400, CAPS pH 10.5, 0.5 M (NH ₄) ₂ SO ₄ , 10% (v/v) glycerol
	D2	40% (v/v) PEG-400, Tris pH 8.5, 0.2 M Li ₂ SO ₄
	D9	20% (v/v) PEG-300, Tris pH 8.5, 5% (w/v) PEG-8000, 10% (v/v) glycerol
	E1	40% (v/v) 2-methyl-2,4-pentanediol cacodylate pH 6.5 5% (w/v) PEG-8000
	G3*	50% (v/v) PEG-200, Tris pH 7.0
	H4	40% (v/v) 2-methyl-2,4-pentanediol, CHES pH 9.5

Table 4.1 (continued) Hit conditions at 15 °C for crystallization of non-heated His₈-tagged *Pfu* DNA polymerase.

Screen name	Well no.	Compositions
The Pegs	A3	0.1 M Sodium acetate pH 4.6, 30 %(v/v) PEG 400
	A4	0.1 M Sodium acetate pH 4.6, 25 %(v/v) PEG 550 MME
	A5	0.1 M Sodium acetate pH 4.6, 25 %(w/v) PEG 1000
	A6	0.1 M Sodium acetate pH 4.6, 25 %(w/v) PEG 2000 MME
	C1	0.1 M Sodium acetate pH 4.6, 25 %(w/v) PEG 3000
	C2	0.1 M Sodium acetate pH 4.6, 25 %(w/v) PEG 4000
	C3	0.1 M Sodium acetate pH 4.6 25 %(w/v) PEG 6000
	C4	0.1 M Sodium acetate pH 4.6, 25 %(w/v) PEG 8000
	C5	0.1 M Sodium acetate pH 4.6, 20 %(w/v) PEG 10000
	C6	0.1 M Sodium acetate pH 4.6, 15 %(w/v) PEG 20000
	G8	0.2 M Magnesium sulfate, 20 %(w/v) PEG 3350
	G11	0.2 M Ammonium sulfate, 20 %(w/v) PEG 3350
	H7	0.2 M Ammonium phosphate, 20 %(w/v) PEG 3350

* Conditions which were considered for crystallization optimization.

4.7 Crystallization of His₈-tagged *Pfu* DNA polymerase

For non-heated His₈-tagged *Pfu* DNA polymerase, the hits of crystallization conditions were shown in table 4.1 (see Figure 1C in APPENDIX C for photographs of crystals). From the table, it was found that the solutions that contained mostly salts and PEG with the molecular weight and concentration range from 200 to 20,000 and 10 to 50%, respectively, gave the best crystals. Due to the beauty of the crystals themselves, which foretell the possibility in giving good diffraction data, availability of chemicals and possibility of self-prepared solutions (solutions that are difficult to prepare might cause irreproducible of the crystals), conditions OZMA 8K10K: G7, OZMA 1K4K: A7, and CryO I/II: G3 (Figure 4.10) were considered for optimization.

Table 4.2 Hit conditions at 15 °C for crystallization of heated His₈-tagged *Pfu* DNA polymerase.

Screen name	Well no.	Compositions
Ozma 8K10K	B1	20% w/v PEG 8000, 200 mM Diammonium tartrate
	B7	20% w/v PEG 8000, 200 mM Lithium sulfate
	C7	20% w/v PEG 8000, 100 mM Potassium phosphate (monobasic)
	E4	10% w/v PEG 10000, 200 mM Ammonium formate
	F3	10% w/v PEG 10000, 200 mM Lithium Acetate
	G9	10% w/v PEG 10000, 200 mM Potassium thiocyanate
	G11	10% w/v PEG 10000, 200 mM Sodium acetate
	H3	10% w/v PEG 10000, 200 mM Sodium formate
	H5	10% w/v PEG 10000, 200 mM Sodium isothiocyanate
Ozma 1K4K	-	-
HTS I	D11	20% w/v PEG 4000, 10% w/v Glycerol, 200 mM Magnesium sulfate
	D5	10% w/v PEG 4000, 10% w/v 2-Propanol, 100 mM Sodium citrate, pH 5.6
	E3	22% w/v PEG 4000, 100 mM Sodium acetate, 200 mM Ammonium sulfate
	E4	25% w/v PEG 4000, 100 mM Sodium citrate, 200 mM Ammonium sulfate, pH 5.6
	F12	8% w/v PEG 8000, 200 mM Lithium Chloride, 50 mM Magnesium sulfate
	G3 *	10% w/v PEG 8000, 100 mM Sodium Acetate, 50 mM Magnesium Acetate
Cryo I/II	D5	25% (v/v) 1,2-propanediol, phosphate-citrate pH 4.2, 5% (w/v) PEG 3000, 10% (v/v) Glycerol
The Pegs	-	-
WIZARDS I/II	B3	10% (w/v) PEG 3000, imidazole pH 8.0, Lithium sulfate
	B7	20% (w/v) PEG 1000, Tris pH 7.0
	C8	0.4 M NaH ₂ PO ₄ / 1.6 M K ₂ HPO ₄ , imidazole pH 8.0, NaCl
	D7	35% (v/v) 2-methyl-2,4-pentanediol, Na/K phosphate pH 6.2
	D9	20% (w/v) PEG 3000, Acetate pH 4.5
	E8	10% (w/v) PEG 8000, Na/K phosphate pH 6.2, NaCl
Hampton HTP	B6	0.2 M Magnesium acetate tetrahydrate, 0.1 M Sodium cacodylate trihydrate pH 6.5, 20% w/v PEG 8000
	F10	0.1 M MES pH 6.5, 12% w/v PEG 20000
HTS II	E2	12% (w/v) Ethanol, 4% (w/v) PEG 400, 100 mM Sodium acetate, pH 4.6

* This condition gave the most single crystals throughout the screening result.

The optimizations were carried out by varying the concentration of the chemical composition of the hit conditions mentioned previously, concentration of protein, and ratio of protein to precipitant. The optimization conditions are shown in APPENDIX D where OPT7, OPT9, and OPT10 are the optimization solutions that were prepared based on the conditions from CryO I/II: G3, OZMA 1K4K: A7, and OZMA 8K10K: G7, respectively. After sitting drop optimization technique at 15 °C, a number of single crystals were obtained from OPT9-F6 which was prepared by diluting the conditions OZMA 1K4K: A7 [30% (w/v) PEG1000, 200 mM Ammonium Phosphate (monobasic)] to 40% using 200 mM ammonium phosphate (monobasic), with a ratio of protein to precipitant of 2 to 1 [protein:precipitant = 2:1], and an initial protein concentration of 15 mg/ml. Figure 4.11 shows a number of single crystals that have been obtained. This condition for crystallization of non-heated His₈-tagged *Pfu* DNA polymerase differed from the report previously published by Goldman, *et al.* (1998).

As for heated His₈-tagged *Pfu* DNA polymerase, the hits of crystallization conditions are shown in table 4.2 (see table 2C in APPENDIX C for the photographs of crystals). From the table, it was found that the solutions suitable for crystallization of heated His₈-tagged *Pfu* DNA polymerase contained salts and high molecular weight PEG ranging from 400 to 10,000 at concentrations ranging from 10 to 25%. Single crystals were obtained from the condition HTS I: G3, which contains 10% w/v PEG8000, 100 mM sodium acetate, and 50 mM magnesium acetate with protein:precipitant = 2:1, starting protein concentrations of 15 and 20 mg/ml, and 15 °C. Please note that after heating the protein, precipitation occurred and caused the protein concentration to decrease 2-3 mg/ml. The precipitated proteins were thought

to be other protein impurities. The crystallization was successfully reproduced with this condition and a number of single crystals were obtained (Figure 4.12).

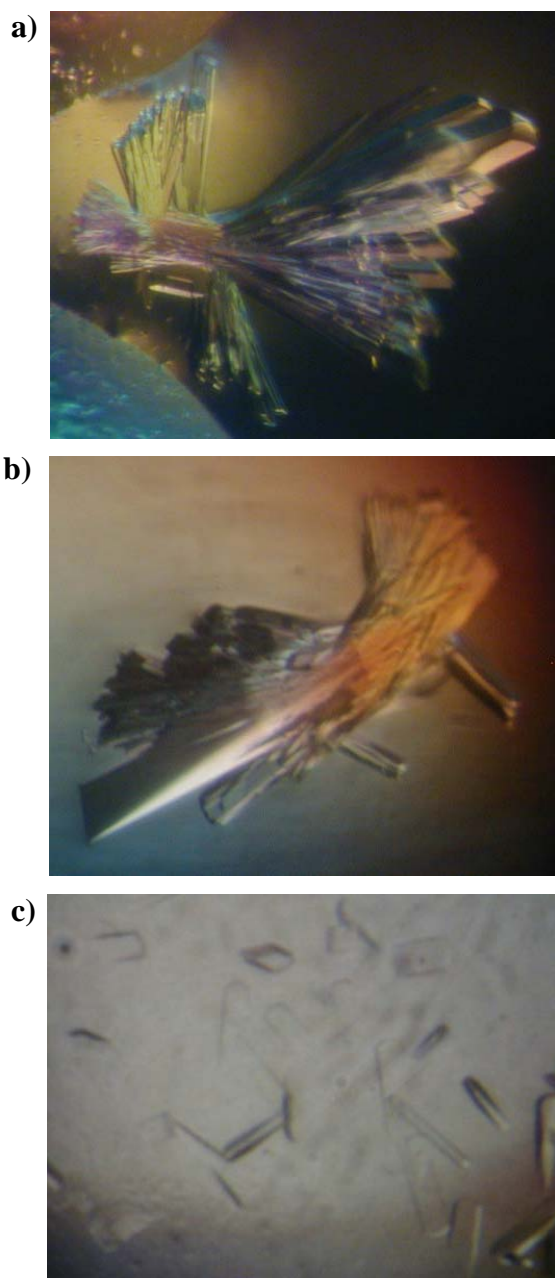


Figure 4.10 Crystals of non-heated His₈-tagged *Pfu* DNA polymerase obtained from the screening plate conditions: **a)** OZMA 8K10K: G7, **b)** OZMA 1K4K: A7 and, **c)** CryO I/II: G3.

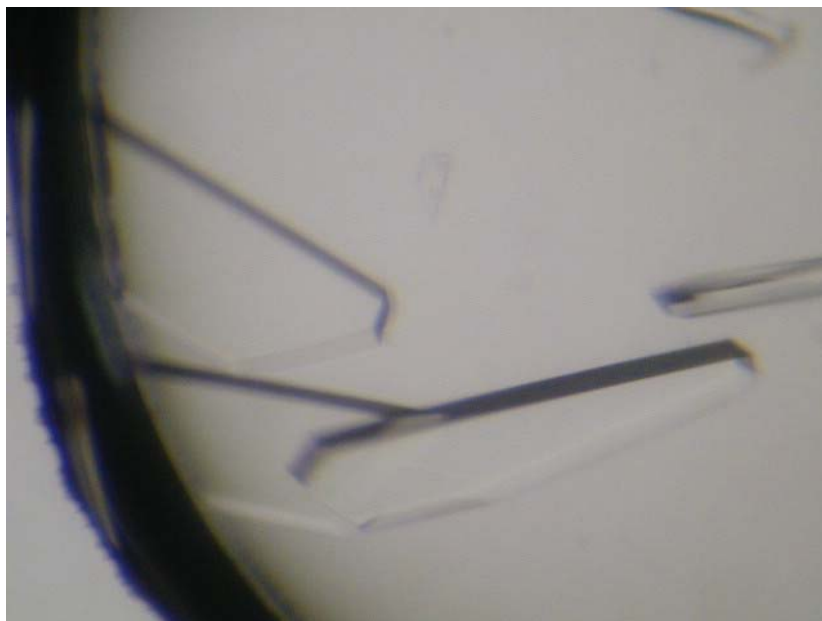


Figure 4.11 Examples of crystals of non-heated His₈-tagged *Pfu* DNA polymerase obtained from optimization OPT9-F6.

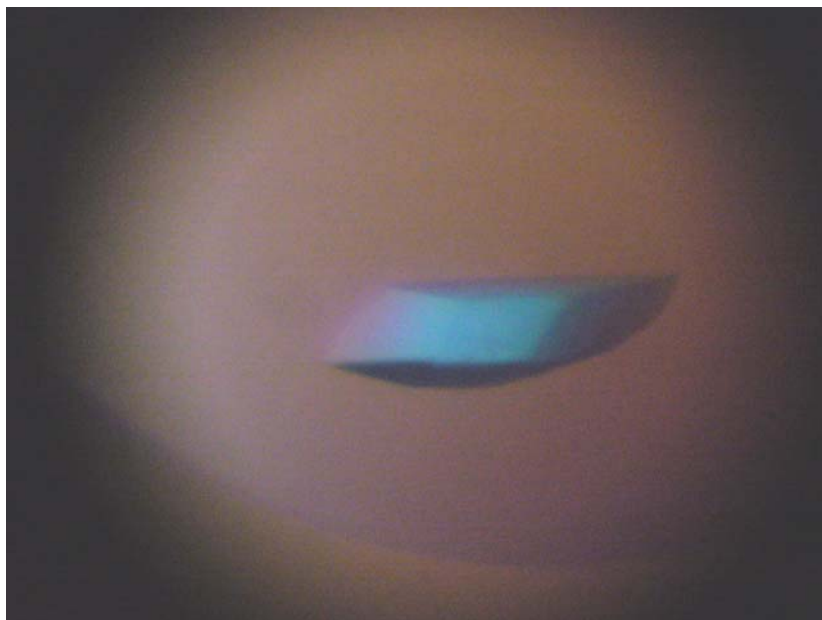


Figure 4.12 Example of crystals of heated His₈-tagged *Pfu* DNA polymerase obtained condition HTS I: G3.

It is possible that heat-treating His₈-tagged *Pfu* DNA polymerase might cause conformation change e.g. from a “closed” conformation to an “opened” conformation or only some loop movement might occur. On the other hand, heating the protein sample might be getting rid of protein impurities which were thought to be precipitated. This might cause a change in the protein sample to be crystallized and affect the crystal growth conditions. Therefore, the growths of crystals of non-heated and heated His₈-tagged *Pfu* DNA polymerase occur in different crystallizing agents.

Table 4.3 Data collection statistics from the heated His₈-tagged *Pfu* DNA polymerase crystal.

Resolution	Reflections ^a	% Completeness ^b	R _{merge} ^c	I/ σ
20.00 - 6.40	1704	89.6	0.058	17.0
6.40 - 5.11	1739	93.1	0.078	15.8
5.11 - 4.47	1763	94.0	0.081	13.3
4.47 - 4.06	1766	95.0	0.105	12.0
4.06 - 3.78	1765	95.3	0.137	11.1
3.78 - 3.55	1772	95.9	0.176	10.2
3.55 - 3.38	1757	94.8	0.239	9.3
3.38 - 3.23	1712	91.8	0.298	8.4
3.23 - 3.11	1597	86.5	0.385	7.1
3.11 - 3.00	1391	75.6	0.427	5.6
20.00 - 3.00	16966	91.2	0.100	12.5

^a Number of unique reflection.

^b Percentage of theoretically possible data measured.

^c R_{merge} ($\sum |I - \langle I \rangle| / \sum \langle I \rangle$)

4.8 Diffraction data

The non-heated His₈-tagged *Pfu* DNA polymerase crystals were able to diffract to a resolution limit of only 4 Å, which is the same as the experiment of Goldman, *et al.* (1998). At this low resolution, the structure could not be solved easily.

The heated His₈-tagged *Pfu* DNA polymerase crystal was able to diffract to a resolution limit of 3 Å. At this resolution of data, the structure could be solved. Unit-cell parameters of the crystal was determined as $a = 91.9$ Å, $b = 126.8$ Å, $c = 88.4$ Å, $\alpha = 90.0^\circ$, $\beta = 109.1^\circ$, and $\gamma = 90.0^\circ$. The crystal is Monoclinic and belongs to the space group C2. The unit-cell parameters of the crystal gave a Matthew's coefficient of $2.64 \text{ Å}^3 \text{Da}^{-1}$ and a solvent content of 53.5% (v/v). The data collection statistics for the heated His₈-tagged *Pfu* DNA polymerase data set are given in table 4.3.

4.9 Preliminary model

The R-factor of 50.9% and free R-factor of 49.3% were obtained after phasing by molecular replacement. The preliminary model, which was obtained after the first round of refinement, has an R-factor of 29.4% and a free R-factor of 36.2% on all 16966 reflections in the range 20.0 Å to 3.0 Å, with an rms $\Delta bonds$ of 0.008 Å, and an rms $\Delta angles$ of 1.152° (symbols defined by XPLOR (Brunger, 1992)). These values can be improved by cycles of model building and refinement. The model consists of residues 24 – 602. There was no interpretable electron density for the N-terminal 23 amino acids – the extra His₈-containing region provided by the pSY5 plasmid vector – which suggested the mobility of this region. The unclear electron density maps occur for residues 603-798.

The overall structure of the heated His₈-tagged *Pfu* DNA polymerase is presented in Figure 4.13. It is basically composed of domains and subdomains, which are the N-terminal domain, Exonuclease domain, and Polymerase domain including the Palm and Fingers subdomains, which is similar to the crystal structures of archaeal family B DNA polymerase that have been solved: *Tgo* DNA polymerase (Hopfner, *et al.*, 1999), Tok DNA polymerase (Zhao, *et al.*, 1999), 9° N-7 DNA polymerase (Rodriguez, *et al.*, 2000), and KOD1 DNA polymerase (Hashimoto, *et al.*, 2001). However, the Thumb domain is absent in the structure due to the unclear electron density maps mentioned earlier, which might be caused by the low resolution of the diffraction data.

Two disulfide bonds are found in the connection site between the Palm and Fingers subdomains of the structure: Cys452-Cys466 and Cys530-Cys533. The electron density maps for these two disulfide bonds are shown in figure 4.14. Hashimoto (2001) proposed that the number of disulfide bonds is correlated with optimum growth temperatures of organisms from the sequence alignment of archaeal DNA polymerases. Hashimoto (2001) also predicted these two disulfide bonds to occur in *Pfu* DNA polymerase. This result, in turn, agrees with the work from Hashimoto (2001).

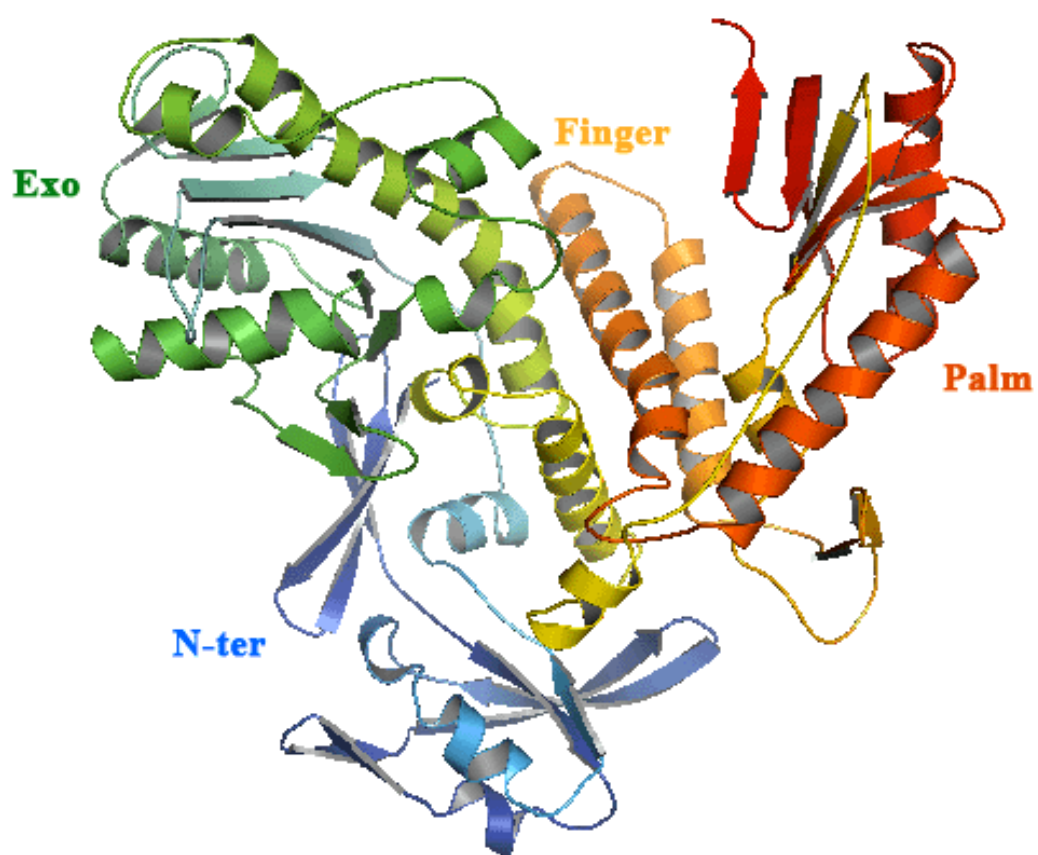


Figure 4.13 Preliminary overall structure of the heated His₈-tagged *Pfu* DNA polymerase. The Figure was prepared using the program PYMOL.

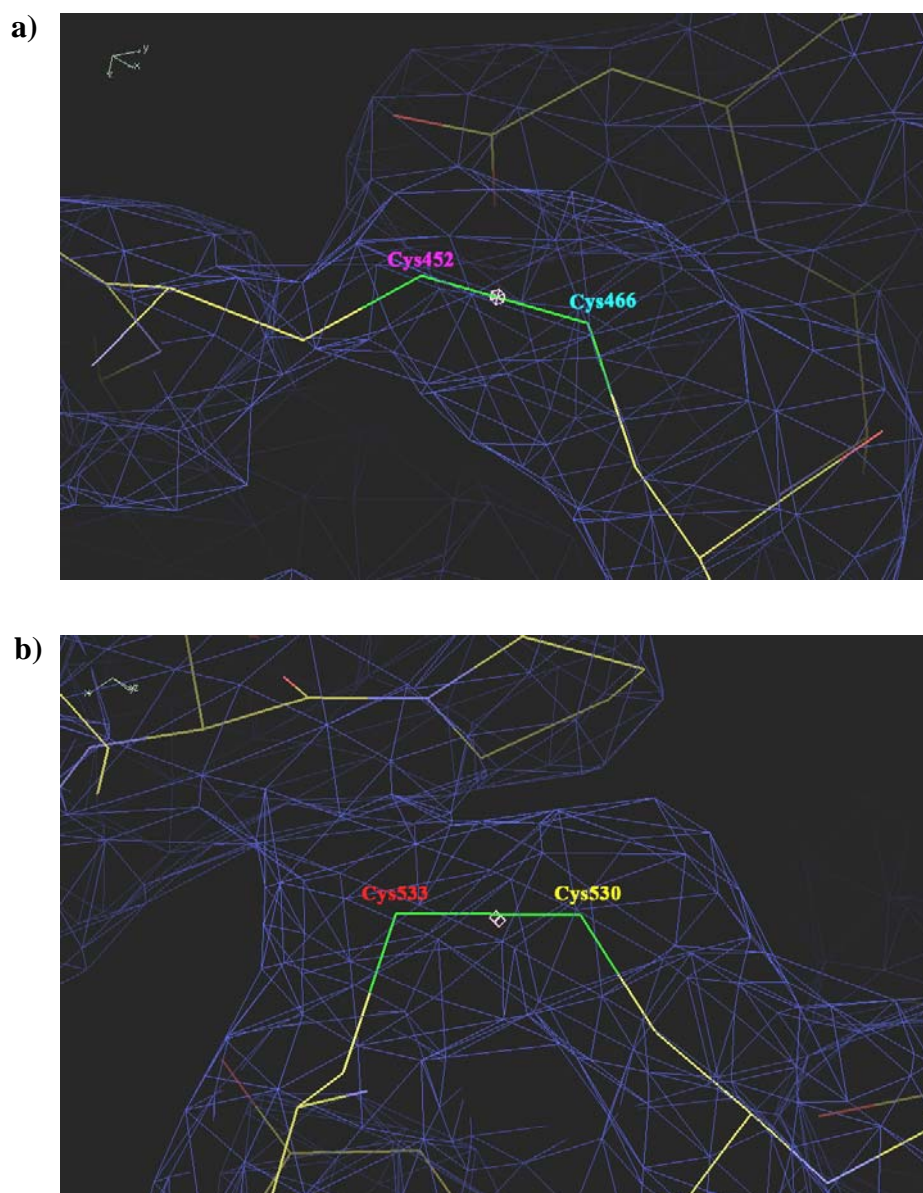


Figure 4.14 The $2|F_o|-|F_c|$ electron density maps contoured at 1σ at 3.0 \AA resolution for the two disulfide bonds in the heated His₈-tagged *Pfu* DNA polymerase. **a)** Cys452-Cys466. **b)** Cys530-Cys533. The Figure was prepared using the program COOT (Emsley and Cowtan, 2004).

CHAPTER V

CONCLUSION

1. The *Pfu* DNA polymerase gene was successfully cloned into the pSY5 expression plasmid after amplification by nested PCR with two sets of primers—set 1 without and set 2 with restriction recognition sites. The cloned gene was 100% identical to the known DNA sequence of *Pfu* DNA polymerase gene.

2. By comparing the small-scale expression from two *E. coli* strains – Rosetta (DE3) pLysS and BL21-Codonplus (DE3)-RIL – the highest expression was obtained from Rosetta (DE3) pLysS at 25 °C, 18 h. This condition was scaled up and the protein yield was 38 mg/L.

3. The expressed recombinant His₈-tagged *Pfu* DNA polymerase was successfully purified by an automated AKTAXpress Purification System. The purified protein was at a satisfactory purity level and suitable for protein crystallization.

4. The relative DNA polymerase activity was determined by comparing the PCR products obtained from His₈-tagged *Pfu* DNA polymerase and the commercial Promega *Pfu* DNA polymerase. The His₈-tagged *Pfu* DNA polymerase was found to have 30,450 U per mg protein.

5. The His₈-tagged *Pfu* DNA polymerase was still stable when incubated at 97.5 °C for 23 h, which was the maximum duration tested.

6. The PCR efficiency of recombinant His₈-tagged *Pfu* DNA polymerase and

the commercial Promega *Pfu* DNA polymerase were compared using λ DNA as the template in amplification of various sizes of amplicons, ranging from 0.5 to 15 kb, with an extension time of 2.5 min. The recombinant His₈-tagged *Pfu* DNA polymerase has higher PCR efficiency than the commercial Promega *Pfu* DNA polymerase. Some loss of PCR efficiency occurred in the heat-treated His₈-tagged *Pfu* DNA polymerases.

7. Single crystals of non-heated His₈-tagged *Pfu* DNA polymerase were obtained from OPT9-F6, which contained 40% of OZMA 1K4K: A7 [30% (w/v) PEG1000, 200 mM ammonium phosphate (monobasic)] diluted using 200 mM ammonium phosphate (monobasic), whereas single crystals of heated His₈-tagged *Pfu* DNA polymerase were obtained from condition HTS I: G3, contained 10% w/v PEG8000, 100 mM sodium acetate, and 50 mM magnesium acetate. Crystallization of both the non-heated and heated enzyme were prepared by protein:precipitant = 2:1 and vapor diffusion-sitting drop method at 15 °C.

8. The non-heated His₈-tagged *Pfu* DNA polymerase crystals were able to diffract to a resolution limit of only 4 Å, whereas the heated His₈-tagged *Pfu* DNA polymerase crystal was able to diffract to a resolution limit of 3 Å. The unit-cell parameters of the heated His₈-tagged *Pfu* DNA polymerase crystal were determined as $a = 91.9$ Å, $b = 126.8$ Å, $c = 88.4$ Å, $\alpha = 90.0^\circ$, $\beta = 109.1^\circ$, and $\gamma = 90.0^\circ$ with the space group of C2. The unit-cell parameters gave a Matthew's coefficient of 2.64 Å³Da⁻¹ and a solvent content of 53.5% (v/v).

9. The preliminary model has an R-factor of 29.4% and a free R-factor of 36.2% with an rms $\Delta bonds$ of 0.008 Å, and an rms $\Delta angles$ of 1.152°. The model consists of residues 24 – 602. The overall structure of the heated His₈-tagged *Pfu*

DNA polymerase is basically composed of an N-terminal domain, Exonuclease domain, and Polymerase domain including the Palm and Fingers subdomains. The Thumb domain is absent in the structure. Two disulfide bonds are found in the connection site between the Palm and Fingers subdomains of the structure: Cys452-Cys466 and Cys530-Cys533.

At this point, the thermostable *Pfu* DNA polymerase has been cloned, sequenced, expressed, and purified for the use in our laboratory with much lower cost and higher efficiency than the Promega commercial *Pfu* DNA polymerase. After some screening and optimization, the crystals of heated and non-heated His₈-tagged *Pfu* DNA polymerase were obtained. The crystals were used for X-ray diffraction data collection. However, the structure of *Pfu* DNA polymerase is yet to be solved.

Suggestion

1. The crystallization optimization could be done with other hit conditions obtained from the crystallization screening results in order to find the crystals that diffract beyond 4 Å of resolution in the case the non-heated His₈-tagged *Pfu* DNA polymerase or 3 Å of resolution in the case of heated His₈-tagged *Pfu* DNA polymerase.

2. In order to study the enzyme-DNA binding, the crystallization in term of complex, co-crystallization, or crystal soaking could be carried out.

3. Intensive comparison of the structure of *Pfu* DNA polymerase in comparison with other DNA polymerase species could be done in order to understand the structural differences related to their function, thermostability, or fidelity.

REFERENCES

REFERENCES

- Bergfors, T. M. (1999). **Protein crystallization: techniques, strategies, and tips**. California: International university line.
- Blow, D. (2002). **Outline of crystallography for biologist**. New York: Oxford university press.
- Blundell, T. and Johnson, L. (1976). **Protein Crystallography**. London: Academic Press.
- Brunger, A. T. (1990). Extension of molecular replacement - a new search strategy based on Patterson correlation refinement. **Acta Crystallogr.** A46: 46-57.
- Brunger, A. T. (1992). Free R-value – a novel statistical quantity for assessing the accuracy of crystal-structures. **Nature**. 355: 472-475.
- Brunger, A. T. and Rice, L. M. (1997). Crystallographic refinement by simulated annealing. **Methods Enzymol.** 277: 243-269.
- CCP4. (1994). The CCP4 suite: Programs for crystallography. **Acta Crystallogr.** D50: 760-763.
- Clegg, W. (1998). **Crystal structure determination**. New York: Oxford university press.
- Cline, J., Braman, C. J. and Hogrefe, H. H. (1996). PCR fidelity of Pfu DNA polymerase and other thermostable DNA polymerases. **Nucleic Acids Res.** 24(18): 3546-3551.
- Cooper, J. (1998). **Molecular replacement** [on-line]. Available: <http://www.soton.ac.uk/~jbc2/molrep/moprep.html>
- Crowther, R. A. and Blow, D. M. (1967). A method of positioning a known molecule

- in an unknown crystal structure. **Acta Crystallogr.** 23: 544-548.
- Dabrowski, S. and Kur, J. (1998). Cloning and expression in *Escherichia coli* of the recombinant His-tagged DNA polymerase from *Pyrococcus furiosus* and *Pyrococcus woesei*. **Protein Expression and Purif.** 14: 131-138.
- Drenth, J. (1994). **Principles of protein X-ray crystallography.** New York: Springer-Verlag.
- Eckert, K. A., and Kunkel, T. A. (1991). DNA polymerase fidelity and the polymerase chain reaction. **PCR Methods Appl.** 1: 17-24.
- Elie, C., de Recondo, A. M., and Forterre, P. (1989). Thermostable DNA polymerase from the archaebacterium *Sulfolobus acidocaldarius*. Purification, characterization and immunological properties. **Eur. J. Biochem.** 178: 619-626.
- Emsley, P. and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. **Acta Crystallogr.** D60: 2126-2132.
- Giacovazzo, C., Monaco, H. L., Viterbo, D., Scordari, F., Gilli, G., Zanotti, G. and Catti, M. (1992). **Fundamentals of crystallography.** New York: Oxford university press.
- Glusker, J. P. and Trueblood, K. N. (1972). **Crystal structure analysis: a primer.** New York: Oxford University Press.
- Goldman, S., Kim, R., Hung, L. W., Jancarik, J. and Kim, S. H. (1998). Purification, crystallization and preliminary X-ray crystallographic analysis of *Pyrococcus furiosus* DNA polymerase. **Acta Crystallogr.** D54: 986-988.
- Hahn, T. (1992). **International Tables of Crystallography. Vol. A: Space-group symmetry.** Boston: Kluwer academic.

- Hammond, C. (1997). **The basics of crystallography and diffraction**. Oxford: Oxford Science Publications.
- Hashimoto, H., Nishioka, M., Fujiwara, S., Takagi, M., Imanaka, T., Inoue, T. and Kai, Y. (2001). Crystal structure of DNA polymerase from hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1. **J Mol Biol.** 306(3):469-477.
- Hirshfield. (1968). Symmetry in the generation of trial structures. **Acta Crystallogr.** A24: 301-311.
- Helliwell, J. R. (1992). **Macromolecular crystallography with synchrotron radiation**. Cambridge: Cambridge university press.
- Hopfner, K. P., Eichinger, A., Engh, R. A., Laue, F., Ankenbauer, W., Huber, R. and Angerer, B. (1999). Crystal structure of a thermostable type B DNA polymerase from *Thermococcus gorgonarius*. **Proc Natl Acad Sci U S A.** 96(7):3600-3605.
- Ito, J., and Braithwaite, D. K. (1991). Compilation and alignment of DNA polymerases. **Nucleic Acids Res.** 19: 4045–4057.
- Jones, T. A., Zou, J. Y., Cowan, S. W. and Kjeldgaard, N. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. **Acta Crystallogr.** A47: 110-119.
- Klimczak, L. J., Grummt, F., and Burger, K. J. (1985). Purification and characterization of DNA polymerase from archaeobacterium *Sulfolobus acidocaldarius*. **Nucleic Acids Res.** 13: 5269–5282.
- Laskowski, R. J., Macarthur, M. W., Moss, D. S. and Thornton, J. M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. **J. Appl. Crystallogr.** 26: 283-290.
- Lattman, E. E. and Love W. E. (1969). A rotational search procedure for detecting a

- known molecule in a crystal. **Acta Crystallogr.** B26: 1854-1857.
- Lattman, E. E. (1972). Optimal sampling of the rotation function. **Acta Crystallogr.** B28: 1065-1068.
- Lu, C. and Erickson, P. H. (1997). Expression in *Escherichia coli* of the thermostable DNA polymerase from *Pyrococcus furiosus*. **Protein Expression and Purif.** 11: 179-184.
- Lundberg, K. S., Shoemaker, D. D., Adams, M. W. W., Short, J. M., Sorge, J. A., and Mathur, E. J. (1991). High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. **Gene.** 108: 1–6.
- Mathur, E. J., Adams, M. W. W., Callen, W. N., and Cline, J. (1991). The DNA polymerase gene from the hyperthermophilic marine archaebacterium, *Pyrococcus furiosus*, shows sequence homology with alpha-like DNA polymerases. **Nucleic Acids Res.** 19: 6952.
- Matthews, B. W. (1968). Solvent content of protein crystals. **J. Mol. Biol.** 33: 491-497.
- Moss, D. S. (1985). The symmetry of the rotation function. **Acta Crystallogr.** A41: 470-475.
- Mroczkowski, B. S., Huvar, A., Lernhardt, W., Misono, K., Nielson, K., and Scott, B. (1994). Secretion of thermostable DNA polymerase using a novel baculovirus vector. **J. Biol. Chem.** 269: 13522–13528.
- Murshudov, G., Vagin, A. and Dodson, E. (1996). Application of Maximum Likelihood Refinement. In **The refinement of protein structures**. Proceedings of Daresbury Study Weekend.
- Murshudov, G. N., Vagin, A. and Dodson, E. J. (1997). Refinement of Macromolecu-

- lar Structures by the Maximum-Likelihood Method. **Acta Cryst.** D53: 240-255.
- Navaza, G. (1994). AMORE: an automated package for molecular replacement. **Acta Crystallogr.** A50: 157-163.
- Otwinowski, Z. (1991). Maximum likelihood refinement of heavy atom parameters. In **Isomorphous replacement and anomalous scattering** (pp 80-86). Daresbury Laboratory, Warrington, UK.
- Otwinowski, Z. (1993). Oscillation data reduction program. In Sawyer, L., Issacs, N. and Bailey, S. (eds.). **Data Collection and Processing** (pp 56-62). SERC Daresbury Laboratory, Warrington, UK.
- Otwinowski, Z. and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In Carter, C. W. Jr. and Sweet, R. M. (eds.). **Methods in enzymology: macromolecular crystallography, part A** vol. 276 (pp 307-326). New York: Academic Press.
- Pannu, N. J., Murshudov, G. N., Dodson, E. J. and Read, R. J. (1998). Incorporation of prior phase information strengthen Maximum-Likelihood structure refinement. **Acta Cryst.** D54: 1285-1294.
- Perler, F. B., Comb, D. G., Jack, W. E., Moran, L. S., Qiang, B., Kucera, R. B., Benner, J., Slatko, B. E., Nwankwo, D. O., Hempstead, S. K., Carlow, C. K. S., and Jannash, H. (1992). Intervening sequences in an Archaea DNA polymerase gene. **Proc. Natl. Acad. Sci. USA** 89: 5577–5581.
- Rella, R., Raia, C. A., Pisani, F. M., D'Auria, S., Nucci, R., Gambacorta, A., de Rosa, M., and Rossi, M. (1990). Purification and properties of a thermophilic and thermostable DNA polymerase from the archaeobacterium *Sulfolobus solfataricus*. **Ital. J. Biochem.** 39: 83–99.

- Rhodes, G. (2006). **Crystallography made crystal clear: a guide for users of macromolecular models** (3rd ed). California: Academic press.
- Rhodes, G. (2006). **Resources for Readers of Crystallography made crystal clear: a guide for users of macromolecular models** [online]. Available: <http://www.usm.maine.edu/~rhodes/CMCC/>
- Rodriguez, A. C., Park, H. W., Mao, C. and Beese, L. S. (2000). Crystal structure of a pol alpha family DNA polymerase from the hyperthermophilic archaeon *Thermococcus sp.* 9 degrees N-7. **J Mol Biol.** 299(2):447-62.
- Stout, G. H. and Jensen, L. H. (1989). **X-ray structure determination—a practical guide** (2nd ed). New York: John Wiley & Sons.
- Uemori, T., Ishino, Y., Toh, H., Asada, K. and Kato, I. (1993). Organization and nucleotide of the DNA polymerase gene from the archaeon *Pyrococcus furiosus*. **Nucleic Acids Res.** 21(2): 259-265.
- Vagin, A. and Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. **J. Appl. Crystallogr.** 30: 1022-1025.
- Wang, J., Sattar, A. K. M. A., Wang, C. C., Karam, J. D., Konigsberg, W. H. and Steiz, T. A. (1997). Crystal structure of a pol a family replication DNA polymerase from bacteriophage RB69. **Cell.** 89: 1087-1099.
- Zhao, Y., Jeruzalmi, D., Moarefi, I., Leighton, L., Lasken, R. and Kuriyan, J. (1999). Crystal structure of an archaebacterial DNA polymerase. **Structure.** 7(10):1189-1199.

APPENDICES

APPENDIX A
SEQUENCE OF CLONED HIS₈-TAGGED
***PFU* DNA POLYMERASE**

atggcagaagaacaccaccaccaccaccaccacctggaagttctgttccaggggcc
M A E E H H H H H H H L E V L F Q G P
gggcggccgatgatttttagatgtggattacataactgaagaaggaaaacctgttattagg
G R P M I L D V D Y I T E E G K P V I R
ctattcaaaaaagagaacggaaaatttaagatagagcatgatagaacttttagaccatac
L F K K E N G K F K I E H D R T F R P Y
atttacgctcttctcagggatgattcaaagattgaagaagttaagaaaataacgggggaa
I Y A L L R D D S K I E E V K K I T G E
aggcatggaaagattgtgagaattgttgatgtagagaaggttgagaaaaagtttctcggc
R H G K I V R I V D V E K V E K K F L G
aagcctattaccgtgtggaaactttatttggaaacatccccaagatgttcccactattaga
K P I T V W K L Y L E H P Q D V P T I R
gaaaaagttagagaacatccagcagttgtggacatcttcgaatacgatatattccatttgca
E K V R E H P A V V D I F E Y D I P F A
aagagatacctcatcgacaaaggcctaataccaatggagggggaagaagagctaaagatt
K R Y L I D K G L I P M E G E E E L K I
cttgcccttcgatatagaaaccctctatcacgaaggagaagagtttggaaaaggcccaatt
L A F D I E T L Y H E G E E F G K G P I
ataatgattagttatgcagatgaaaatgaagcaaaggtgattacttggaaaaacatagat
I M I S Y A D E N E A K V I T W K N I D
cttcacatcggttgaggttgatatcaagcgagagagagatgataaagagatttctcaggatt
L P Y V E V V S S E R E M I K R F L R I
atcagggagaaggatcctgacattatagttacttataatggagactcattcgacttccca
I R E K D P D I I V T Y N G D S F D F P
tatttagcgaaaaagggcagaaaaacttgggattaaattaaccattggaagagatggaagc
Y L A K R A E K L G I K L T I G R D G S
gagcccaagatgcagagaataggcgatatgacggctgtagaagtcaagggaagaatacat
E P K M Q R I G D M T A V E V K G R I H
ttcgacttgatatcatgtaataacaaggacaataaatctcccaacatacacactagaggct
F D L Y H V I T R T I N L P T Y T L E A
gtatatgaagcaatttttggaaagccaaaggagaaggtatacgccgacgagatagcaaaa
V Y E A I F G K P K E K V Y A D E I A K
gcctgggaaagtggagagaaccttgagagagttgccaaatactcgatggaagatgcaaag
A W E S G E N L E R V A K Y S M E D A K
gcaacttatgaactcgggaaagaatttccttccaatggaaattcagctttcaagattagtt
A T Y E L G K E F L P M E I Q L S R L V
ggacaacctttatgggatgtttcaaggtcaagcacaggggaacctttagagtggttctta
G Q P L W D V S R S S T G N L V E W F L
cttaggaaagcctacgaaagaaacgaagtagctccaaacaagccaagtgaagaggagtat
L R K A Y E R N E V A P N K P S E E E Y
caaagaaggctcagggagagctacacaggtggattcgttaaagagccagaaaaggggttg
Q R R L R E S Y T G G F V K E P E K G L
tgggaaaacatagtatacctagatttttagagccctatatccctcgattataattaccac
W E N I V Y L D F R A L Y P S I I I T H
aatgtttctcccgatactctaaatcttgagggatgcaagaactatgatatcgctcctcaa
N V S P D T L N L E G C K N Y D I A P Q
gtaggccacaagttctgcaaggacatccctggttttataccaagtctcttgggacatttg
V G H K F C K D I P G F I P S L L G H L
ttagaggaaagacaaaagattaagacaaaaatgaaggaaactcaagatcctatagaaaaa
L E E R Q K I K T K M K E T Q D P I E K

```

atactccttgactatagacaaaaagcgataaaactccttagcaaattctttctacggatat
I L L D Y R Q K A I K L L A N S F Y G Y
tatggctatgcaaaagcaagatggtagtgtaaggagtgtgctgagagcggttactgcctgg
Y G Y A K A R W Y C K E C A E S V T A W
ggaagaaagtacatcgagtttagtatggaaggagctcgaagaaaagtttggaatttaaagtc
G R K Y I E L V W K E L E E K F G F K V
ctctacattgacactgatggctctctatgcaactatcccaggaggagaaagtgaggaaata
L Y I D T D G L Y A T I P G G E S E E I
aagaaaaaggctctagaatttgtaaaatacataaattcaaagctccctggactgctagag
K K K A L E F V K Y I N S K L P G L L E
cttgaatatgaagggtttttataagaggggattccttcgttacgaagaagaggtatgcagta
L E Y E G F Y K R G F F V T K K R Y A V
atagatgaagaaggaaaagtcattactcgtgggttttagagatagttaggagagattggagt
I D E E G K V I T R G L E I V R R D W S
gaaattgcaaaagaaactcaagctagagttttggagacaataactaaaacacggagatggtt
E I A K E T Q A R V L E T I L K H G D V
gaagaagctgtgagaatagtaaaagaagtaatacaaaagcttgccaattatgaaattcca
E E A V R I V K E V I Q K L A N Y E I P
ccagagaagctcgcaatatatgagcagataacaagaccattacatgagtataaggcgata
P E K L A I Y E Q I T R P L H E Y K A I
ggtcctcacgtagctgttgcaaagaaactagctgctaaaggagttaaaataaagccagga
G P H V A V A K K L A A K G V K I K P G
atggtaattggatacatagtagtagggcgatgggtccaattagcaatagggaatttcta
M V I G Y I V L R G D G P I S N R A I L
gctgaggaatacgcgatccccaaaagcacaagtatgacgcagaatattacattgagaaccag
A E E Y D P K K H K Y D A E Y Y I E N Q
gttcttcagcggtacttaggatattggaggggatttggtacagaaaggaagacctcaga
V L P A V L R I L E G F G Y R K E D L R
taccaaaagacaagacaagtcggcctaacttcctggcttaacattaaaaaatcctag
Y Q K T R Q V G L T S W L N I K K S *

```

Figure 1A Sequence of cloned His₈-tagged *Pfu* DNA polymerase. The highlighted region is the extra nucleotides or amino acids which were added into the original *Pfu* DNA polymerase sequence after cloning into the pSY5 plasmid vector.

* indicates Stop codon.

APPENDIX B

DETERMINATION OF RELATIVE POLYMERASE

ACTIVITY

Calculation of relative polymerase activity of His₈-tagged *Pfu* DNA polymerase

From figure C.1, the ratio of band intensity in lane 2 and lane 8 equals:

$$\frac{734.440}{361.870} = 2.03 \approx 2$$

Then, 0.2 µg of the His₈-tagged *Pfu* DNA polymerase possesses:

$$2 \times 3 \text{ U} = 6 \text{ U}$$

Thus, 1 mg of the His₈-tagged *Pfu* DNA polymerase contains:

$$\left| \frac{6 \text{ U}}{0.2 \text{ µg}} \right| \left| \frac{1 \text{ µg}}{10^{-3} \text{ mg}} \right| = 30,000 \text{ U}$$

Consequently, the relative polymerase activity of the His₈-tagged *Pfu* DNA polymerase is 30,000 U per milligram protein.

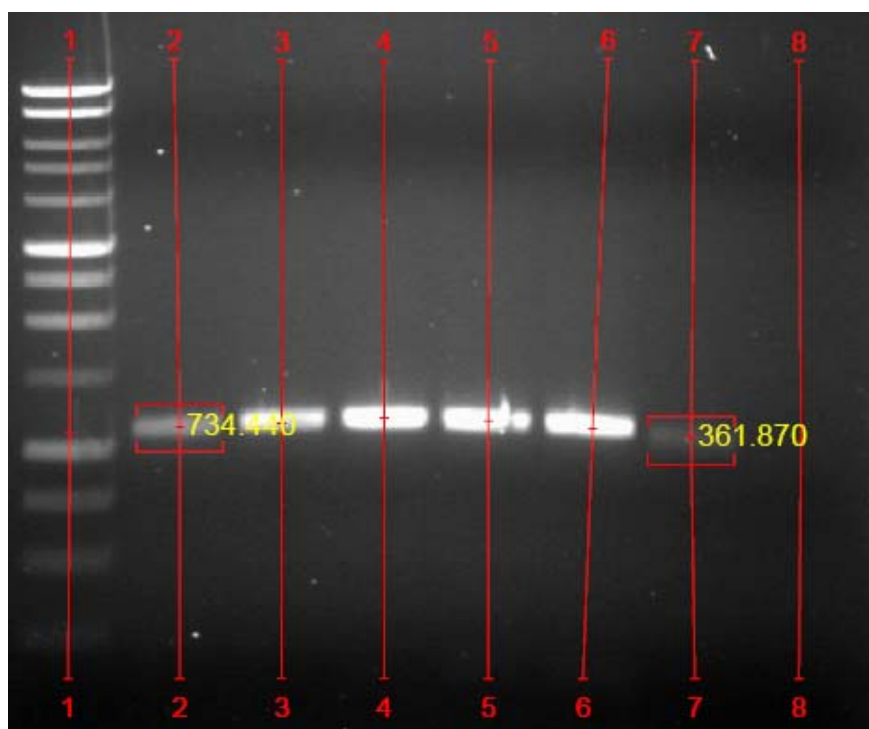


Figure 1B The band intensities determined by Quantity One® Software, with the arbitrary unit. The numbers in lanes 2 and 8 indicate the intensities of the bands when 0.2 μg of the His₈-tagged *Pfu* DNA polymerase and 3 U of the commercial Promega *Pfu* DNA polymerase were used, respectively.

APPENDIX C

EXAMPLE OF CRYSTALS FROM CRYSTALLIZATION

SCREENING



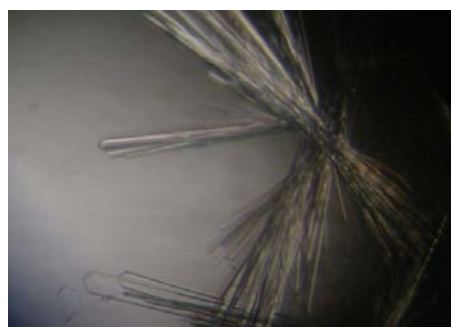
The Pegs: A5



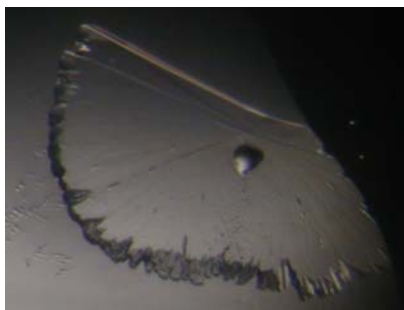
OZMA 1K4K: H9



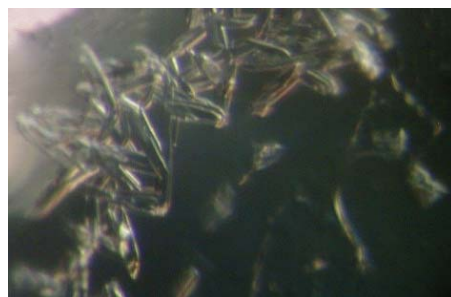
OZMA 1K4K: E11



Cryo I/II: A1



The Pegs: A4



OZMA 1K4K: E12

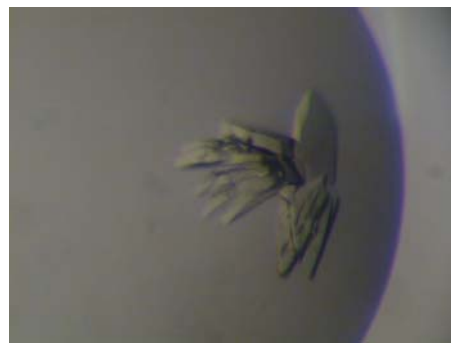


OZMA 8K10K: B4

Figure 1C Example of crystals of non-heated His₈-tagged *Pfu* DNA polymerase obtained from various conditions of crystallization screening.



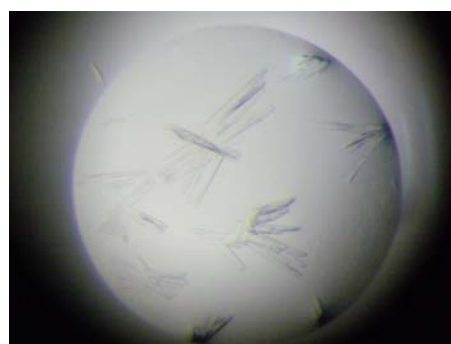
OZMA 8K10K: F3



OZMA 8K10K: G11



HTS I: F12



HTS I: D11



CryO I/II: D5



OZMA 8K10K: G11

Figure 2C Example of crystals of heated His₈-tagged *Pfu* DNA polymerase obtained from various conditions of crystallization screening.

APPENDIX D

CRYSTALLIZATION OPTIMIZATION FORMULATION

Table 1D Crystallization optimization condition OPT7 (based on condition CryO I/II: G3).

Name	Formulation
OPT7-F1	50 % PEG200, 100 mM Tris HCl, pH 6.8
OPT7-F2	50 % PEG200, 100 mM Tris HCl, pH 6.9
OPT7-F3	50 % PEG200, 100 mM Tris HCl, pH 7.0
OPT7-F4	50 % PEG200, 100 mM Tris HCl, pH 7.1
OPT7-F5	50 % PEG200, 100 mM Tris HCl, pH 7.2
OPT7-F6	50 % PEG200, 100 mM Tris HCl, pH 7.3
OPT7-F7	45 % PEG200, 100 mM Tris HCl, pH 6.8
OPT7-F8	45 % PEG200, 100 mM Tris HCl, pH 6.9
OPT7-F9	45 % PEG200, 100 mM Tris HCl, pH 7.0
OPT7-F10	45 % PEG200, 100 mM Tris HCl, pH 7.1
OPT7-F11	45 % PEG200, 100 mM Tris HCl, pH 7.2
OPT7-F12	45 % PEG200, 100 mM Tris HCl, pH 7.3
OPT7-F13	35 % PEG200, 100 mM Tris HCl, pH 6.8
OPT7-F14	35 % PEG200, 100 mM Tris HCl, pH 6.9
OPT7-F15	35 % PEG200, 100 mM Tris HCl, pH 7.0
OPT7-F16	35 % PEG200, 100 mM Tris HCl, pH 7.1
OPT7-F17	35 % PEG200, 100 mM Tris HCl, pH 7.2
OPT7-F18	35 % PEG200, 100 mM Tris HCl, pH 7.3

Table 2D Crystallization optimization condition OPT9 (based on condition Ozma 1K4K: A7).

Name	Formulation
OPT9-F1	90 % of OPT9 ^a with 200 mM ammonium phosphate (monobasic)
OPT9-F2	80 % of OPT9 ^a with 200 mM ammonium phosphate (monobasic)
OPT9-F3	70 % of OPT9 ^a with 200 mM ammonium phosphate (monobasic)
OPT9-F4	60 % of OPT9 ^a with 200 mM ammonium phosphate (monobasic)
OPT9-F5	50 % of OPT9 ^a with 200 mM ammonium phosphate (monobasic)
OPT9-F6	40 % of OPT9 ^a with 200 mM ammonium phosphate (monobasic)
OPT9-F7	80 % of OPT9 ^a with water
OPT9-F8	70 % of OPT9 ^a with water
OPT9-F9	60 % of OPT9 ^a with water
OPT9-F10	50 % of OPT9 ^a with water
OPT9-F11	40 % of OPT9 ^a with water
OPT9-F12	90 % of OPT9 ^a with water

^aOPT9 contains 30 % PEG1000 and 200 mM ammonium phosphate (monobasic)

Table 3D Crystallization optimization condition OPT9 (based on condition Ozma 8K10K: G7).

Name	Formulation
OPT10-F1	90 % of OPT10 ^b with 50 mM potassium phosphate (monobasic)
OPT10-F2	80 % of OPT10 ^b with 50 mM potassium phosphate (monobasic)
OPT10-F3	70 % of OPT10 ^b with 50 mM potassium phosphate (monobasic)
OPT10-F4	60 % of OPT10 ^b with 50 mM potassium phosphate (monobasic)
OPT10-F5	50 % of OPT10 ^b with 50 mM potassium phosphate (monobasic)
OPT10-F6	40 % of OPT10 ^b with 50 mM potassium phosphate (monobasic)

^bOPT10 contains 10 % PEG10,000 and 50 mM potassium phosphate (monobasic)

APPENDIX E

PROGRAMS USED IN THIS STUDY

List of programs used in this study

AMORE	Molecular replacement package (Navaza, 1994).
CAD	Utility program which combines MTZ files and outputs data in the same portion or reciprocal space (CCP4, 1994).
COOT	A stand-alone portion of CCP4's Molecular Graphics project. Its focus is crystallographic model-building and manipulation rather than representation.
DENZO	Zbyszek Otwinowski's interactive film and image plate processing package. Auto indexing proceeds as in REFIX but on a single image. The input parameters are refined by a least squares method minimizing the discrepancy between the centroids of the spots and the predicted diffraction spot positions. A 3-dimensional profile is calculated based on profiles learned from other reflections to calculate a profile fitted intensity. Also takes care of Lorentz, polarization and air absorption effects (Otwinowski, 1993).
FFT	Crystallographic fast Fourier transformation - calculates electron density maps and Patterson maps from reflection data using the fast Fourier transform algorithm. FFT is space group specific for P_1 and

EXPAND must be run to prepare the data for this space group. The method breaks up the 3D FFT into a number of 1D FFTs, via a process called Beevers-Lipson factorization. These 1D FFTs can then be calculated rapidly. The program was used here to calculate Fouriers, difference Fouriers, Pattersons and difference Pattersons (CCP4, 1994).

HKL2000 A package of programs intended for the analysis of X-ray diffraction data collected from single crystals, and consists of three programs: *XDISPLAYF* for visualization of the diffraction pattern, *DENZO* for data reduction and integration, and *SCALEPACK* for merging and scaling of the intensities obtained by *DENZO* or other programs.

MOLREP An automated program for molecular replacement (Vagin and Teplyakov, 1997).

O General model building, model manipulation and graphics display program (Jones, *et al.*, 1991). It is part of the CCP4 package.

PROCHECK Program to check the stereochemistry of a model (Laskowski, *et al.*, 1993)

PYMOL An open-source tool to visualize molecules available from (www.pymol.org). It has excellent capabilities in creating high-quality

images from 3D structures, well developed functions for manipulating structures and some basic functions to analyze their chemical properties.

REFMAC Macromolecular refinement program. The program can carry out rigid body, [tls](#), restrained or unrestrained refinement against X-ray data, or idealisation of a macromolecular structure. It minimises the coordinate parameters to satisfy either a Maximum Likelihood or Least Squares residual.

SCALEPACK Package used to scale and merge the output from DENZO. The output is a unique set of intensities. The program uses profile fitting and rejects spots with bad agreement, adds partials, calculates scale factors and B-factors for each batch of data and can perform post refinement (Otwinowski, 1991).

SFALL Structure factor calculations using FFT (CCP4, 1994).

SHELL_SCALE Scaling program that allows for scaling in resolution shells, used here for scaling F_{calc} to F_{obs} (Stuart, 1989).

SHELX Program used here for Patterson map interpretation. Selects peaks from Patterson and interprets them using a vector superposition approach which reduces the number of peaks to be interpreted

(Sheldrick *et al.*, 1993).

SIGMAA Method of phase combination, usually for isomorphous and calculated phases. (Read, 1986).

XDISPLAYF Program used for visualization and measurement of the diffraction pattern.

The Molecular Replacement method is described below (Cooper, 1998).

The Rotation Search - carried out to determine the correct orientation(s) of the model by maximizing the overlap between the stationary Patterson map calculated from the measured intensities and the rotated interpolated values of the model Patterson map. This rotation function search selects intramolecular vectors by restricting the choice of vectors in the observed Patterson to a relatively small radius of Patterson space (75-80% of the molecular diameter) and producing the calculated Patterson in a PI cell of approximately twice the size of the molecule. Peaks close to the origin are removed as they are dominated by the Patterson origin peak. Orientations are ranked in order by the value of the rotation, a simplified form of correlation coefficient.

Patterson Correlation Refinement - is used as a gauge to substantiate the solutions from the rotation function and to achieve some refinement of the solution. This refinement can be carried out by minimizing the following function for each promising rotation function solution:

$$E_{TOTAL} = 1 - \frac{\langle |E_{obs}|^2 |E_m|^2 - \langle |E_{obs}|^2 \rangle \langle |E_m|^2 \rangle \rangle}{\sqrt{\langle |E_{obs}|^4 - \langle |E_{obs}|^2 \rangle^2 \rangle \langle |E_m|^4 - \langle |E_m|^2 \rangle^2 \rangle}}$$

E_{obs} and E_m denote the observed and the model calculated normalized structure factors respectively. The angle brackets denote averaging over the set of reflections expanded to P_1 . A correct orientation should correspond to the highest value of the target function.

The Translation Function - is used to superimpose the intermolecular vectors by translating the rotated model through either real or Patterson space. In this study the optimal translation was assessed by maximizing the agreement between the measured and model calculated normalized square structure factors (E^2).

$$E_{TOTAL} = W_a \left(1 - \frac{\langle |E_{obs}^2|^2 |E_m^2|^2 - \langle |E_{obs}^2|^2 \rangle \langle |E_m^2|^2 \rangle \rangle}{\sqrt{\langle |E_{obs}^2|^4 - \langle |E_{obs}^2|^2 \rangle^2 \rangle \langle |E_m^2|^4 - \langle |E_m^2|^2 \rangle^2 \rangle}} \right)$$

Where W_a is a weighting function. The search need only be carried out over the "Cheshire Cell" as defined by Hirshfield (1968). Unfortunately the translation search often proves to be the most problematic stage of molecular replacement.

Positional Refinement - XPLOR implements restrained refinement to increase the number of observations by minimizing:

$$E_{total} = E_{empirical} + E_{effective}$$

Where $E_{\text{empirical}}$ is the empirical energy term given by:

$$E_{\text{empirical}} = E_{\text{bonds}} + E_{\text{angles}} + E_{\text{dihedral}} + E_{\text{impropers}} + E_{\text{nonbonded}}$$

The four conformational energy terms are described by harmonic restraints about idealized values of bond lengths and angles to help preserve the good chemical sense of the model. The nonbonded term takes into consideration van der Waals and electrostatic interactions both within and between molecules to prevent bad contacts. $E_{\text{effective}}$ is described by a number of energy terms that use experimental information and restrained parameters. The X-ray term is given by:

$$E_{\text{X-ray}} = \frac{W_A}{N_A} \sum_h w_h [|F|_{\text{obs}}(h) - k |F|_{\text{calc}}(h)]^2$$

where W_a and W_h are the overall weight and the weight on a reflection respectively, N_A is a normalization constant and k is a scaling constant between the observed and calculated reflections. Minimization is carried out by conventional conjugate gradient refinement which often gets trapped in local minima. Positional refinement can be combined with molecular dynamics and simulated annealing to overcome these false minima.

Molecular Dynamics Refinement - Here the atoms are mathematically heated and assigned initial velocities provided by a Maxwellian distribution to overcome local energy minima. The molecule is then cooled (simulated annealing) to allow the molecule to achieve the lowest energy state compatible with the restraints provided by $E_{\text{effective}}$. The process involves solving Newton's equation of motion for all atoms.

$$m_i \frac{\delta^2 r_i}{\delta t^2} = -\nabla E_{\text{total}}$$

APPENDIX F

CRYSTALLOGRAPHIC THEORY

Protein crystal forms

Proteins theoretically are able to adopt a total of 65 possible crystal forms. These arise from the seven distinct crystal systems; primitive, triclinic, monoclinic, orthorhombic, tetragonal, hexagonal and cubic lattices. Centering lattices extends this list to 14, termed the Bravais Lattices. Non-translational elements are defined as operations which applied repeatedly to a point within the system to bring it back to its original position and orientation - a set of such elements are called a point group, and may comprise of combinations of mirror planes, rotation axes and centres of inversion. For the seven crystal systems there are 32 different combinations of rotations and mirror planes with translational elements, termed screw axis and glide plane respectively, which when taken with the 14 Bravais Lattices result in a total of 230 separate arrangements or space groups. This number is somewhat reduced in the case of protein crystals to 65, due to the chirality of proteins (L-amino acids) which does not permit crystallization involving mirror or glide planes or centers of inversion. Further details are documented in the International Tables of Crystallography (volume A), edited by Hahn (1992).

Darwin's equation

The total energy ($E_{(hkl)}$) of a reflection from ideal crystal rotating at a constant angular velocity (ω) is given by:

$$E_{(hkl)} = \frac{e^4}{m^2 c^4 \omega} I_0 LPA \frac{V_x}{V_0^2} |\mathbf{F}_{(hkl)}|^2$$

where: L is the Lorentz factor, P is the polarization correction, A is the absorption correction, V_x is the volume of the crystal, V_o is the volume of the unit cell, I_0 is the incident intensity and $|\mathbf{F}_{(hkl)}|$ the structure factor for reflection hkl .

Table 1H Protein crystal symmetry

Crystal system	Unit cell dimension	Point group	Space groups
Triclinic	$a \neq b \neq c; \alpha \neq \beta \neq \gamma \neq 90^\circ$	1	P1
Monoclinic	$a \neq b \neq c; \alpha = \gamma = 90^\circ; \beta > 90^\circ$	2	P2, P2 ₁ C2
Orthorhombic	$a \neq b \neq c; \alpha = \beta = \gamma = 90^\circ$	222	P222, P222 ₁ , P2 ₁ 2 ₁ 2, P2 ₁ 2 ₁ 2 ₁ C222, C222 ₁ I222, I2 ₁ 2 ₁ 2 ₁ F222
Tetragonal	$a = b \neq c; \alpha = \beta = \gamma = 90^\circ$	4	P4, P4 ₁ , P4 ₃ , P4 ₂
		422	P422, P42 ₁ 2, P4 ₁ 22, P4 ₃ 22, P4 ₁ 2 ₁ 2, P4 ₃ 2 ₁ 2, P4 ₂ 22, P4 ₂ 2 ₁ 2
		4	I4, I4 ₁
		422	I422, I4 ₁ 22
Trigonal	$a = b = c; \alpha = \beta = \gamma \neq 90^\circ$	3	P3, P3 ₁ , P3 ₂ , R3
		32	P312, P321, P3 ₁ 12, P3 ₁ 21, P3 ₂ 12, P3 ₂ 21, R32
Hexagonal	$a = b = c; \alpha = \beta = 90^\circ; \gamma = 120^\circ$	6	P6, P6 ₁ , P6 ₂ , P6 ₃ , P6 ₄ , P6 ₅
		622	P622, P6 ₁ 22, P6 ₂ 22, P6 ₃ 22, P6 ₄ 22, P6 ₅ 22
Cubic	$a = b = c; \alpha = \beta = \gamma = 90^\circ$	23	P23, P2 ₁ 3
		432	P432, P4 ₁ 32, P4 ₂ 32, P4 ₃ 32
		23	I23, I2 ₁ 3
		432	I432, I4 ₁ 32
		23	F23
		432	F432, F4 ₁ 32

Maximum angle of rotation

The maximum oscillation of a crystal possible before overlapped reflections occur can be calculated by the following equation.

$$\Delta_{rot} = \tan^{-1}(d_{min} / cell\ edge) - spot\ width$$

Where: Δ_{rot} = the maximum oscillation and d_{min} = min resolution. *Spot width* would typically be 0.2° . Often the maximum oscillation is not appropriate for actual data collection because of the trade off between efficiency of data collection and increased background.

Bragg's Law

Bragg's Law describes the conditions for constructive interference when the incident and diffracted rays are in the same plane as the normal to the diffracted rays.

$$n\lambda = 2d\sin\theta$$

Where n is an integer, θ is the angle of incidence and reflection and d is the distance between lattice planes. This equation can be used to calculate the unit cell dimensions from an image of a principal zone. The cell parameters (d) in two directions can be determined from the spacing between the spots (Δ) using the following equation.

$$d = \frac{\lambda}{2 \sin(\tan^{-1}(\Delta / s) / 2)}$$

and the third dimension from the spacing of the rings arising from different layers in reciprocal space where r is the radius of the n th circle.

$$d = \frac{n\lambda}{1 - \cos(\tan^{-1}(r/s))}$$

R_{merge}

R_{merge} is a measure of the accuracy of scaled data. Where I_{hi} is the intensity of a reflection and I_h is the average intensity of that reflection.

$$R_{\text{merge}}(I) = \frac{\sum_h \sum_i |I_h - I_{hi}|}{\sum_h \sum_i I_{hi}}$$

Structure factor equation

Describes the scattering from a crystal over all atoms (j):

$$F_{(hkl)} = \sum_{j=1}^N f_j \exp\left[\frac{-B \sin^2 \theta}{\lambda^2}\right] \exp 2\pi i(lx_j + ky_j + lz_j)$$

Where: hkl are integers (the reciprocal lattice), f are atomic scattering factors and B is the temperature factor that models the thermal movement of the atoms:

$$B = \frac{8\pi^2 \bar{u}^2}{3}$$

Where \bar{u}^2 is the square of the mean displacement along the normal to the reflecting planes. A large B-factor results in rapid fall off of intensity with θ . Where atoms are not free to vibrate equally in all directions the isotropic B-factor can be replaced by an anisotropic expression:

$$\exp - (b_{11}h^2 + b_{12}hk + b_{13}hl + b_{22}k^2 + b_{23}kl + b_{33}l^2)$$

The phase problem arises from the structure factor equation being a complex quantity

$$\mathbf{F}_{(\mathbf{hkl})} = F_{(hkl)} \exp i\alpha_{(hkl)}$$

where $F_{(hkl)}$ is the amplitude and $\alpha_{(hkl)}$ are the phases. As the measured quantity is the intensity (i.e. $|\mathbf{F}_{(\mathbf{hkl})}|^2$) the phases are lost.

Electron density equation

The following equation relates the electron density (ρ) at any point to the structure factors.

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l \mathbf{F}_{(\mathbf{hkl})} \exp(-2\pi i(hx + ky + lz))$$

Where V is the volume of the unit cell.

Patterson synthesis

The Patterson function, $P(xyz)$, can be calculated from the structure factor amplitudes, and represents a map of vectors between atoms.

$$P(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l |\mathbf{F}_{(\mathbf{hkl})}|^2 \cos(-2\pi i(hu + kv + lw))$$

Where u , v , and w are coordinates in Patterson space.

BIBLIOGRAPHY

Mr. Anantasak Loonchanta was born in June 7th, 1980 in Khon Kaen, Thailand. He graduated with the Bachelor of Engineering in Chemical Engineering, Suranaree University of Technology (SUT), Nakhon Ratchasima, Thailand in 2003. During the last year of his bachelor degree study, he has an opportunity to work as a research assistant on molecular cloning in the School of Biotechnology, Institute of Agricultural Technology in the same university. He achieved his first master degree in Master of Science in Chemistry at the Department of Chemistry in collaboration with Institute of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden in 2005. The work in his first master degree was mainly protein X-ray crystallography. After master degree graduation, he came back to the School of Biotechnology, SUT and became a research assistant on DNA finger printing. Whilst working as a research assistant, he has received the SUT scholarship to continue his second master degree in the School of Biotechnology, SUT. Recently, he has been accepted for a PhD position in the Department of Chemistry, University of British Columbia, Vancouver, Canada.